

Université de Montréal

**Analyse cognitive de la cohérence interindexeurs  
lors de l'indexation de documents**

par  
Claire David

Département de Communication  
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)  
en Sciences de la communication

Juin 2003

© Claire David 2003



P  
90  
U51  
2004  
1.007

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

**Analyse cognitive de la cohérence interindexeurs  
lors de l'indexation de documents**

présentée par :  
Claire David

a été évaluée par un jury composé des personnes suivantes :

Aude Dufresne  
présidente-rapporteur

Luc Giroux  
directeur de recherche

Louis-Claude Paquin  
membre du jury

John E. Leide  
examineur externe

Suzanne Bertrand-Gastaldy  
représentante du doyen de la FES

## Résumé

L'indexation consiste à analyser un document et lui assigner une liste de termes (mots-clés, descripteurs, etc.) représentant son contenu. Cette liste est intégrée à une base de données et sert au repérage du document. Les recherches sur l'indexation ont démontré que les indexeurs diffèrent quant au choix de termes d'indexation pour décrire le contenu d'un document, ce qui entraîne un repérage souvent inefficace, remettant en question la fiabilité des systèmes d'information.

La cohérence interindexeurs, définie comme la capacité chez plusieurs indexeurs de produire des listes de termes similaires pour représenter un même document, a fait l'objet de nombreuses études qui n'ont cependant pu fournir d'explications satisfaisantes quant à la nature et à la provenance des variations rencontrées. Pour mieux comprendre ce manque de cohérence, il semble judicieux d'examiner de près les processus intellectuels sous-jacents à l'activité d'indexation.

Étant donné la nature intellectuelle de la tâche et les diverses contraintes qui y sont associées, notre approche consiste à analyser le travail intellectuel des indexeurs sous un nouvel angle, celui des théories de la résolution de problème provenant des sciences cognitives. On cherche alors à repérer les causes des variations interindexeurs par la mise à jour des stratégies intellectuelles utilisées.

Pour ce faire, trois types complémentaires de verbalisations sont utilisés: 1) verbalisation concomitante recueillie pendant l'exécution de la tâche, 2) verbalisation consécutive sur trace qui consiste à recueillir les commentaires et réponses de l'indexeur lorsqu'il est confronté à l'enregistrement vidéo de son propre comportement, 3) entrevue lors de laquelle l'indexeur commente les listes de descripteurs produites par différents indexeurs pour les mêmes documents.

Les données ont été recueillies lors d'une expérimentation au site normal de travail; quatre indexeurs experts d'un centre de documentation spécialisé ont indexé les mêmes quatre documents à l'aide de deux thésaurus.

Il est démontré que dans ce contexte, la mise à jour d'une représentation personnelle de la solution permet de rendre compte d'une bonne part des variations interindexeurs. Nos analyses ont permis de dégager différents profils d'indexeurs basés sur l'utilisation hiérarchisée d'un ensemble d'éléments de connaissance particulier (usager, précision, actualité, etc.). Ainsi, ce qui constitue une indexation *satisfaisante* est un composite de deux, trois ou quatre éléments de connaissance privilégiés et pondérés différemment selon l'indexeur et résulte de la prise en compte simultanée des multiples contraintes de la tâche.

Nous croyons que ces résultats pourraient être utiles à la formation des indexeurs. En effet, ces derniers pourraient être sensibilisés à l'impact de la prise en compte des divers éléments de connaissance sur leurs choix de termes et descripteurs. Cette conscientisation n'entraînerait pas nécessairement une amélioration de la cohérence interindexeurs, mais pourrait faire ressortir l'importance d'établir des rencontres régulières pour les indexeurs d'un même centre afin qu'ils puissent mettre en commun leurs stratégies individuelles et qu'en conséquence, ils puissent ajuster leurs processus d'indexation. Ainsi, nous supposons que de tels échanges permettraient probablement une meilleure cohérence, non pour les choix précis de termes, mais pour l'application des divers éléments de connaissance.

Mots-clés:

Indexation, cohérence interindexeurs, résolution de problème, processus cognitifs, protocoles verbaux, satisficing, compréhension de texte

## Abstract

Indexing is the process by which an indexer analyzes a document and assigns a list of keywords that represents its content. However, it is well known that indexers differ significantly in their judgment as to which set of terms reflects most adequately the content of a document, thereby limiting the efficiency of future searching.

Interindexer consistency, defined as the ability for several indexers to assign similar lists of keywords to a document, has been studied for decades. These studies have identified several factors associated with the lack of interindexer consistency, but failed to provide a comprehensive model of this phenomenon. We argue that in order to understand the low level of interindexer consistency, it is necessary to study the processes underlying indexing activity.

Given that indexing is first and foremost an intellectual activity, one has to borrow from theories and methods of cognitive psychology. Therefore, we use a different perspective to study interindexer consistency by considering indexing as a problem solving situation. Since we want to uncover intellectual activities involved in indexing in order to grasp causes of variations in the choices of keywords, we need to access indexers' mental processes.

In order to access these cognitive processes, three kinds of verbal reports are used: 1) thinking aloud protocol collected while the task is in progress, 2) retrospective comments provided by the indexer when confronted with a video recording of his own behavior, 3) indexer's comments on other lists of keywords produced for the same document.

We will present results of an experiment in which four expert indexers were asked to index the same four documents with two different thesauri.

It will be shown that in this context, indexers' personal representation of the solution provides explanation for most of interindexer variations. Our analysis also bring out different profiles of indexers based on a hierarchical use of different clusters of criteria

(user's needs, precision, novelty, etc.). Thus, a *satisfying* indexing can be described as a composite made of two, three or four preferred criteria which are weighted differently by each indexer. This is also viewed as the outcome of an ongoing negotiation process between the various task's constraints and the application of preferred criteria.

We believe that these results could be useful for indexers training. Indexers could be made aware of the consequences of possible use of different kinds of clusters of criteria and their impact on the choice of keywords. This awareness would not necessarily improve the level of interindexer consistency. But, indexers working in a same documentary center could benefit from regular meetings where each would get insights from others on the indexing strategy put forward in particular situations. In return, each indexer could adjust his(her) indexing process accordingly, and this might lead to a greater consistency in uses of criteria.

Keywords:

Indexing, interindexer consistency, problem solving, cognitive processes, verbal protocols, satisficing, text comprehension

## Table des matières

Résumé	iii
Abstract	v
Liste des tableaux	xi
Liste des figures	xiii
Dédicace	xiv
Remerciements	xv
<b>INTRODUCTION GÉNÉRALE</b>	1
<b>CHAPITRE PREMIER : Contexte théorique</b>	6
<b>1. Indexation de documents écrits</b>	8
1.1 Chaîne documentaire et repérage de l'information	9
1.2 Langages documentaires	12
1.3 Description de la tâche d'indexation	19
1.4 Critères de qualité	21
1.5 Problématique de la cohérence interindexeurs	23
1.6 La nature du sujet et sa représentation	30
<b>2. Approche cognitive pour l'étude de l'indexation</b>	38
2.1 Compréhension de texte : Modèles de van Dijk et Kintsch	42
2.1.1 <i>Microstructure et macrostructure</i>	43
2.1.2 <i>Macrorègles</i>	45
2.1.3 <i>Macrostratégies</i>	48
<u>Macrostratégies contextuelles</u>	49
<u>Macrostratégies textuelles</u>	51
2.2 Recherches sur l'indexation en tant que compréhension de texte	55
2.2.1 <i>Beghtol</i>	55
2.2.2 <i>Farrow</i>	57
2.3 Indexation en tant que situation de résolution de problème	66
2.3.1 <i>Description de la résolution de problème</i>	67
2.3.2 <i>Stratégies d'indexation</i>	68
2.3.3 <i>Problèmes bien et mal définis</i>	74
2.3.4 <i>Indexation comme problème mal défini</i>	77

2.3.5 <i>Application des éléments de connaissance</i>	81
<u>Exhaustivité</u>	83
<u>Spécificité</u>	87
<u>Exactitude</u>	89
<u>Pertinence</u>	90
<u>Usager</u>	93
<u>Approches orientées document et usager</u>	95
<b>CHAPITRE II: MÉTHODOLOGIE</b>	99
<b>1. Analyse de protocoles verbaux</b>	103
1.1 Verbalisation concomitante	104
1.2 Verbalisation consécutive sur trace	107
1.3 Évaluation comparée des indexats	110
1.4 Transcription, segmentation et codage des protocoles	112
1.5 Élaboration des catégories	113
<b>2. Situation expérimentale</b>	115
2.1 Choix des sujets	116
2.1.1 <i>Profil des indexeurs</i>	117
2.2 Choix des documents	118
2.3 Choix des langages documentaires	119
2.4 Déroulement de l'expérimentation	121
2.4.1 <i>Description de la tâche demandée</i>	121
2.4.2 <i>Production de listes</i>	122
2.4.3 <i>Réalisation de la tâche</i>	123
2.4.4 <i>Entraînement à la verbalisation</i>	124
2.5 Recueil des protocoles concomitants	124
2.5.1 <i>Condition particulière</i>	125
2.6 Recueil des protocoles consécutifs	126
2.7 Évaluation comparée des indexats	128
<b>3. Données recueillies</b>	129
3.1 Listes	130
3.2 Protocoles	130
3.3 Évaluation comparée des indexats	130
3.4 Matériel audiovisuel	131

3.5 Dispositif expérimental	131
<b>4. Transcriptions des protocoles recueillis</b>	<b>132</b>
4.1 Segmentation des protocoles et catégories	133
<b>5. Calculs de cohérence</b>	<b>134</b>
<b>CHAPITRE III: RÉSULTATS</b>	<b>138</b>
<b>1. Analyses quantitatives</b>	<b>138</b>
1.1 Identification des indexeurs	139
1.2 Présentation des documents	139
1.3 Taux de cohérence interindexeurs	141
1.4 Listes de descripteurs	143
1.5 Caractérisation des indexeurs	150
<b>2. Analyses qualitatives</b>	<b>153</b>
2.1 Déroulement de la procédure d'indexation	154
2.1.1 <i>Utilisation et connaissance du thésaurus Envirodoq</i>	158
2.1.2 <i>Révision de l'indexation</i>	165
2.1.3 <i>Utilisation du thésaurus Infoterra</i>	166
2.1.4 <i>Utilisation des éléments textuels</i>	171
2.2 Repérage de l'utilisation des éléments de connaissance	174
2.3 Prise en compte des éléments de connaissance par les indexeurs	181
2.3.1 <i>Exhaustivité</i>	182
2.3.2 <i>Précision et exactitude</i>	188
2.3.3 <i>Usager</i>	197
2.3.4 <i>Éléments de connaissance supplémentaires</i>	201
2.4 Comparaison interindexeurs pour l'utilisation des éléments de connaissance	212
2.5 Hiérarchisation de l'utilisation des éléments de connaissance	218
2.5.1 <i>E1</i>	218
2.5.2 <i>E2</i>	223
2.5.3 <i>E3</i>	230
2.5.4 <i>E4</i>	234
2.6 Évaluation comparée des indexats	240
2.6.1 <i>Listes de descripteurs modifiées et taux de cohérence interindexeurs</i>	240
2.6.2 <i>Commentaires des indexeurs sur l'ensemble des indexats</i>	243
2.7 Portraits cognitifs des quatre indexeurs	248

	x
DISCUSSION GÉNÉRALE	254
CONCLUSION	270
Références	273
Annexe A : Listes de descripteurs produites par les indexeurs ( <u>feuilles détachables</u> )	xvi
Annexe B : Feuilles brouillon et listes de descripteurs produites par les indexeurs ( <u>feuilles détachables</u> )	xxi
Annexe C : Protocoles concomitant et consécutif de l'indexation du document <i>Étang aéré</i> par E3	xxvi
Annexe D : Sections consultées du document <i>Étang aéré</i>	xlvi
Annexe E : Sections consultées du document <i>Biogaz</i>	liii
Annexe F : Sections consultées du document <i>Compostage</i>	lxiv
Annexe G : Sections consultées du document <i>Énergie</i>	lxxi

## Liste des tableaux

Tableau I	Extrait de la Classification Décimale Universelle (CDU)	14
Tableau II	Facteurs potentiels de variation de la cohérence interindexeurs	25
Tableau III	Principaux résultats des études sur les facteurs affectant la cohérence interindexeurs	26
Tableau IV	Interrelation entre les représentations du sujet, les types d'information et les méthodes d'indexation	37
Tableau V	Informations contextuelles utilisées pour les macrostratégies	50
Tableau VI	Caractéristiques des thésaurus Envirodoq et Infoterra	120
Tableau VII	Liste des protocoles et entrevue recueillis pour un indexeur associés à la durée de chaque étape	130
Tableau VIII	Taux de cohérence interindexeurs pour les quatre documents	141
Tableau IX	Taux de cohérence interindexeurs pour les quatre documents en fonction de divers ensembles de descripteurs	142
Tableau X	Listes de descripteurs produites pour le document <i>Étude de sites d'implantation d'étangs aérés</i>	144
Tableau XI	Listes de descripteurs produites pour le document <i>Élimination du biogaz. Sélection d'une technologie et évaluation des impacts</i>	145
Tableau XII	Listes des descripteurs produites pour le document <i>Compostage des boues résiduaires urbaines</i>	146
Tableau XIII	Listes des descripteurs produites pour le document <i>Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal</i>	147
Tableau XIV	Taux de cohérence de chaque indexeur par rapport au groupe pour l'indexation des quatre documents	151
Tableau XV	Condensé de l'indexation du document <i>Étang aéré</i> et identification partielle de la procédure d'indexation et des éléments de connaissance utilisés par E3	176
Tableau XVI	Ensemble des éléments de connaissance privilégiés par chaque indexeur	212

Tableau XVII	Modification des listes de descripteurs produites pour le document <i>Étude de sites d'implantation d'étangs aérés</i>	241
Tableau XVIII	Modification des listes de descripteurs produites pour le document <i>Élimination du biogaz. Sélection d'une technologie et évaluation des impacts</i>	242
Tableau XIX	Taux de cohérence interindexeurs pour la modification des listes de descripteurs des documents <i>Étang aéré</i> et <i>Biogaz</i>	243
Tableau XX	Hierarchie d'importance des éléments de connaissance pour chaque indexeur	248

## Liste des figures

Figure 1	Chaîne documentaire et repérage de l'information : Rôle de l'indexation	10
Figure 2	Tour de Hanoï	74
Figure 3	Organigramme du processus d'indexation utilisant un thésaurus	79

À Gaston...

## Remerciements

Luc, ce fut un immense privilège d'étudier avec toi en tant que directeur de thèse. Avoir la certitude qu'une personne est constamment à l'écoute et toujours disponible est un trésor inestimable dans cette traversée solitaire. Tes conseils, tes critiques et tes encouragements ont servi de phare à mon gouvernail. Nous avons fait un long, très long parcours ensemble et durant ces années, j'ai pu apprécier ton enthousiasme débordant, ta vive intelligence et ton intégrité. J'ai appris beaucoup de toi, Luc et les acquis que me lègue notre collaboration vont bien au-delà d'une thèse. Merci pour ton amitié. Merci pour tout.

Je tiens également à remercier Suzanne Bertrand-Gastaldy, professeure à l'École de bibliothéconomie et des sciences de l'information, qui, à bien des égards, a agit officieusement en tant que co-directrice pour une grande partie de ce travail. Suzanne, je suis profondément reconnaissante pour tout le soutien que tu m'as apporté.

Mes sincères remerciements à tous les professeur(e)s du Département de communication qui ont participé à ma formation du 1<sup>er</sup> cycle jusqu'aux séminaires doctoraux, ainsi qu'à mes collègues et ami(e) étudiant(e)s qui m'ont accompagnée dans ce périple. Je remercie également l'ensemble du personnel de soutien.

Je remercie tous les indexeurs qui ont accepté si spontanément de participer à l'expérimentation. Cette thèse n'aurait pu prendre forme sans leur contribution.

Merci à Jacques et Denis qui m'ont assistée en tant que techniciens pour les enregistrements vidéo et audio de l'expérimentation.

Gros merci à mon équipe de correctrices, Anne, Claire, Danielle, Josette, Marie-Josée et Sylvie.

Merci à mes parents, Denise et Roger, d'avoir toujours été là pour moi et de m'avoir fourni les outils pour découvrir et aimer la vie. Une partie de cette thèse vous revient.

Merci à mes deux sœurs de m'avoir si "taquinement" encouragée. Anne, merci pour la transcription des protocoles. Now, do you know that *we've* got the greatest thing in the world?

Katéri, la venue au monde de Morgan et Éloïse m'a permis de reconnaître l'importance réelle des choses. Merci.

Merci à toutes les copines et à tous les copains. L'espace ici ne me permet pas d'écrire chacun de vos noms, mais vous vous reconnaîtrez, j'en suis certaine.

À ma grande compagne d'armes, Florence. Que de discussions passionnantes et "vraies", autant sur la science que sur la vie! Merci pour ta précieuse amitié.

Denis, merci d'être là. Merci de croire en moi. Sans ton amour, tout ceci n'aurait pas de sens.

## INTRODUCTION GÉNÉRALE

L'indexation consiste à représenter le contenu d'un document par l'assignation de termes (mots-clés, vedettes-matière, descripteurs, etc.) issus d'un langage documentaire (répertoire de vedettes-matière, thésaurus, etc.) ou du langage naturel permettant une recherche d'information par sujets contrairement à une recherche à partir d'éléments bibliographiques connus, tels auteurs, édition, année de parution, etc. L'expansion rapide de l'Internet, le nombre croissant de bases de données et le développement des bibliothèques virtuelles accentuent la nécessité d'une indexation de qualité, tout en soulevant la complexité grandissante et les difficultés liées à une telle entreprise (Bates, 1998; Fidel, 1994; Lancaster, 1998; Lunin, 1994).

L'efficacité d'un système d'information ne dépend pas uniquement de la qualité de l'indexation. Les problèmes de repérage proviennent également d'une imprécision dans l'interrogation du système en rapport avec le degré d'expertise de l'utilisateur dans le domaine de sa recherche et dans l'interrogation, le travail du bibliothécaire-consultant, le langage d'interrogation, la logique de recherche sous-jacente aux outils de recherche (Bates, 2002; Bertrand, 1993). Cependant, une mauvaise indexation constitue la cause principale des problèmes liés au repérage (Chaumier, 1988).

L'intérêt actuel en science de l'information (SI) est au développement des méthodes d'indexation automatisée, mais pour plusieurs chercheurs (Milstead, 1994, 2000; Suchman, 1995; Todd 1992), il est primordial de centrer les efforts de recherche sur le travail intellectuel des indexeurs. Milstead (1994) rapporte que les systèmes d'indexation automatisée actuels utilisent des méthodes statistiques, syntaxiques et sémantiques qui tentent d'imiter les résultats produits par l'indexation humaine mais leur succès demeure relativement limité étant donné le manque de connaissance des opérations intellectuelles sous-jacentes au travail des indexeurs.

Toutefois, l'indexation humaine reste privilégiée dans la plupart des centres de documentation, incluant les centres spécialisés, puisque l'indexeur humain parvient à une compréhension globale que ne peuvent atteindre les systèmes automatisés (Albrechtsen, 1993; Lancaster, 2003; Milstead, 1994, 2000).

L'indexation humaine n'est cependant pas infaillible. En effet, il est reconnu en SI que plusieurs indexeurs diffèrent pour le choix de termes représentant le contenu d'un document, résultant en un repérage inadéquat et surtout incomplet, créant une incertitude constante pour l'utilisateur face à l'efficacité du repérage (Rolling, 1981). Cet accord interindexeurs ou cohérence interindexeurs se rapporte à "la capacité chez plusieurs indexeurs d'analyser un même document de la même façon et de représenter les mêmes concepts par les mêmes termes" (Bertrand-Gastaldy, 1986, p.11).

Les études sur la cohérence interindexeurs ont permis d'identifier une trentaine de facteurs associés aux variations de choix de termes, tels que l'expertise en indexation, la connaissance du domaine, l'utilisation d'un vocabulaire contrôlé (Leonard, 1977; Markey, 1984). Cependant, ces études, de nature strictement expérimentale, ne permettent pas de fournir un modèle explicatif au constat d'incohérence interindexeurs. Curieusement, les vingt dernières années fournissent très peu de références concernant cette problématique (Bertrand, 1993; Lancaster, 1998; Sievert et Andrews, 1991).

Il semble donc pertinent de réexaminer le travail intellectuel des indexeurs afin de fournir une meilleure explication au constat d'incohérence interindexeurs et ainsi parvenir à une meilleure compréhension de cette tâche fondamentale pour le fonctionnement des systèmes d'information.

Pour ce faire, nous examinons les processus cognitifs des indexeurs à l'aide des théories et méthodes des sciences cognitives. D'un point de vue théorique, l'indexation sera considérée comme une situation de résolution de problème (Newell et Simon, 1972) et plus particulièrement comme un problème mal défini (c'est-à-dire ne comportant pas de solution unique) pour lequel la solution est déterminée en partie par le solutionneur, en l'occurrence l'indexeur (Simon, 1973).

En effet, les instructions fournies par la norme internationale ISO 5963 de l'Organisation internationale de normalisation (ISO, 1985), ainsi que les normes nationales française NF Z 47-102 de l'Association française de normalisation (AFNOR, 1978) et anglaise BS 6529 de la British Standard Institution (BSI, 1984) et même celles de plusieurs

politiques institutionnelles sont de nature relativement générale et laissent une place importante au jugement de l'indexeur. Par exemple, on indique à l'indexeur que "le nombre de descripteurs attribués à un document ... doit être déterminé entièrement par la quantité d'information contenue dans le document en liaison avec les besoins supposés des utilisateurs" (ISO, 1985, p.3), lesquels, en réalité sont rarement identifiables avec précision. Ou encore, avec l'utilisation d'un thésaurus, on suggère "de sélectionner le terme le plus spécifique existant pour représenter une notion donnée" (ISO, 1985, p.4), ce qui devient problématique lorsqu'un concept du document n'a pas d'équivalent dans le thésaurus.

Nous supposons que l'incohérence relative entre indexeurs experts prend sa source dans la représentation personnelle de chaque indexeur en ce qui concerne les composantes mal définies du problème de l'indexation.

En ce sens, nous posons l'hypothèse générale qu'avec l'expérience, l'indexeur expert a assimilé les normes et règles de l'indexation de manière personnelle, et en conséquence a développé sa propre méthode pour produire une indexation "satisfaisante".

En d'autres termes, qu'est-ce qui permet à l'indexeur de déterminer que son indexation est complétée? Puisqu'il doit se fier à son propre jugement, sur quoi base-t-il l'évaluation de ses choix de termes? Qu'est-ce qu'un terme "indexable" pour l'indexeur expert? Est-ce que des indexeurs experts de profil similaire utilisent les mêmes critères?

Pour répondre à ces questions, il s'agit d'examiner de près les divers éléments de connaissance utilisés par les indexeurs experts afin d'avoir un portrait précis de ce que doit comporter une "bonne" indexation (Bell et Jones, 1979) ou une solution satisfaisante pour chaque indexeur. En retour, la systématisation de l'utilisation de ces éléments de connaissance devrait permettre d'expliquer la variation de choix de termes.

Puisque l'on suppose que le manque de cohérence interindexeurs peut être causé par une appréhension et une application différentes des éléments de connaissance utilisés pour la résolution de l'indexation (exhaustivité, précision, exactitude, usager, etc.), on doit

---

<sup>1</sup>Le terme satisfaisante renvoie à la notion de *satisficing* développée par Simon (1973) sur laquelle nous reviendrons tout au long de la présente recherche.

pouvoir repérer et caractériser l'utilisation de ces divers éléments qui permettent à un indexeur de déterminer que son indexation est satisfaisante. Ainsi, pour vérifier notre hypothèse générale, nous avons choisi d'étudier le travail intellectuel réel d'indexeurs experts en procédant à une expérimentation qui permet de recueillir des données directement en cours de tâche.

En somme, l'approche choisie devrait permettre, non seulement d'expliquer le phénomène de la cohérence interindexeurs d'une façon nouvelle, mais également d'expliquer certains aspects des processus intellectuels impliqués pour l'indexation qui n'ont reçu que très peu d'attention. Par ailleurs, l'utilisation combinée de plusieurs types de verbalisations apparaît comme un moyen efficace de faire ressortir les différentes traces du travail intellectuel des indexeurs.

La présente recherche est divisée en cinq chapitres. Le premier chapitre, consacré au contexte théorique, comprend deux parties. La première partie présente de manière détaillée la tâche de l'indexeur et le contexte dans lequel l'indexation s'effectue afin de mettre en perspective l'importance de son rôle pour le fonctionnement des systèmes d'information. Les études antérieures se rapportant au problème de la cohérence interindexeurs sont présentées ainsi que l'ambiguïté entourant la nature et la représentation du sujet d'un document. La deuxième partie présente l'approche cognitive pour l'étude de l'indexation. Ainsi, les recherches qui ont examiné la tâche d'indexation sous l'angle des sciences cognitives sont présentées, ce qui permet de situer notre étude en rapport avec elles et de montrer en quoi notre perspective s'en inspire et s'en distingue.

Le deuxième chapitre présente la méthodologie utilisée pour le recueil et le traitement des données. Dans un premier temps, la méthode d'analyse de protocoles verbaux est décrite en termes généraux afin de démontrer l'apport de chaque type de verbalisation utilisé. Ensuite, les différents éléments de la situation expérimentale sont présentés ainsi que la procédure de traitement.

Le troisième chapitre présente les résultats obtenus à l'aide de la méthode utilisée. Une première partie rapporte les résultats des analyses quantitatives. En plus de fournir les taux de cohérence interindexeurs, ces analyses permettent une première comparaison

interindexeurs à partir des listes produites. La deuxième partie, consacrée aux analyses qualitatives, représente le cœur de cette recherche. Ces analyses exposent la manière dont chaque indexeur utilise les divers éléments de connaissance et ce que la comparaison de ces prises en compte permet d'expliquer au niveau des variations interindexeurs. Ces analyses démontrent l'influence prédominante de certains éléments ou combinaison d'éléments sur l'évaluation des choix de termes et descripteurs de chaque indexeur. Finalement, nous exposons ce que cette hiérarchisation permet d'expliquer au niveau des variations interindexeurs et la manière dont elle définit les conditions d'une indexation satisfaisante propres à chaque indexeur.

Finalement, la discussion et la conclusion abordent les résultats en termes plus généraux afin de faire ressortir ce qu'ils permettent d'expliquer face à la problématique du manque de cohérence interindexeurs et ce qu'ils mettent en lumière sur la tâche d'indexation ainsi que la contribution de cette recherche et les implications pour des travaux futurs.

## CHAPITRE PREMIER : Cadre théorique

*J'ai pu trouver ce que je cherchais  
parce que je suis monté sur les épaules  
de la génération qui m'avait précédé.*

Sir Isaac Newton

Les technologies de l'information et de la communication (TIC) constituent un élément incontournable à l'intérieur de plusieurs secteurs de l'activité humaine (scientifique, social, politique, culturel, etc.). Qu'il s'agisse d'études portant sur les usages des TIC et les environnements sociocognitifs où ils prennent place (Bardini, 1996; Hutchins, 1995; Millerand, 2003; Suchman, 1987, 1993), le développement de nouveaux environnements informatiques (Dufresne, 2002) ou la gestion de l'information pour en optimiser l'accès et l'utilisation (Dupourqué, 2002; Saàdani et Bertrand-Gastaldy, 2000), les phénomènes liés aux TIC sont examinés sous divers angles par des chercheurs provenant d'un vaste éventail de disciplines, œuvrant conjointement à maintes occasions<sup>1</sup>.

Cette réalité se reflète également à travers la nécessité d'une collaboration interdisciplinaire pour l'amélioration de l'ensemble des éléments constituant les systèmes d'information afin de garantir un repérage efficace de l'information.

En effet, l'un des aspects fondamentaux des TIC repose sur la capacité et l'efficacité du traitement et du repérage de l'information, et ce, plus particulièrement depuis l'avènement d'Internet, comme le souligne d'ailleurs Lancaster (1998), "The Internet has caused a tremendous growth of interest in information retrieval techniques in general and automatic methods in particular" (p.277).

Les outils de repérage sont de plus en plus sophistiqués et puissants, mais cette puissance est d'une utilité limitée si elle ne peut bénéficier des savoirs et savoir-faire d'experts de tous les domaines impliqués pour la réalisation d'un système vraiment efficace.

---

<sup>1</sup> À ce titre, mentionnons le Centre de recherche interdisciplinaire sur les technologies émergentes (CITÉ) de l'Université de Montréal fondé par les départements des sciences de la communication, d'informatique et recherche opérationnelle, et par l'École de bibliothéconomie et sciences de l'information. Le CITÉ, dont les recherches portent principalement sur les usages innovants des TIC, s'est assuré la collaboration de spécialistes provenant de domaines aussi variés que la psychologie, la sociologie, le droit, la linguistique, la santé, le marketing, les mathématiques, la démographie et l'éducation.

En effet, tous les éléments d'un système d'information sont interreliés et comportent des implications pour les autres éléments du système: la structure de la base de données, les métadonnées, l'indexation, les moteurs de recherche, les capacités de recherche, la conception de l'interface, les comportements de recherche des usagers, etc. Par exemple, un repérage sera infructueux ou pauvre, si la logique sous-jacente au langage d'interrogation n'est pas adaptée à la logique sous-jacente au langage documentaire utilisé pour la classification ou l'indexation<sup>2</sup> (Bates 2002). Comme le mentionne Bates (2002) à propos des bibliothèques virtuelles :

Digital libraries cannot be fully effective as information sources for users until the entire design process is done in a manner that involves genuine conceptual and practical coordination among the people working on the system layers. The information content, its database structure, and retrievable elements, should not be selected without full consultation with experts in the subject domain and in the information seeking behavior and context of use of the proposed digital library information. (p.397)

Plus près de nos intérêts de recherche, l'indexation constitue un mode de classification qui permet de représenter le contenu des documents sous forme de liste de descripteurs. Généralement, il s'agit d'une analyse pointue effectuée avec un vocabulaire spécialisé se rapportant à un domaine particulier de connaissances.

Tout comme les autres éléments d'un système d'information, l'indexation se doit d'être de qualité puisque une mauvaise indexation aura une incidence directe sur la qualité du repérage. "Variability in depth of indexing and in quality of indexing can have an enormous influence on quality and power of results at the point of actual searching" (Bates, 2002, p.387).

Ainsi, le contexte actuel des TIC fait ressortir l'importance de bien comprendre le travail des indexeurs, non seulement pour assurer l'efficacité lors du repérage, mais également parce que c'est à partir de la connaissance de cette tâche que sont élaborées les

---

<sup>2</sup> Bates (2002) donne plusieurs exemples concrets de problèmes qui surgissent lorsqu'un ou plusieurs éléments du système a été élaboré sans tenir compte des paramètres ou caractéristiques précises des autres éléments.

méthodes d'indexation automatique. Actuellement toutefois, les recherches sur l'indexation se rapportent en majorité au développement de méthodes d'extraction et d'assignation automatique de mots-clés plutôt qu'à la compréhension du travail intellectuel des indexeurs.

Bien que les procédures mises au point soient relativement efficaces pour assister les indexeurs dans leur tâche, aucune ne parvient à un degré de performance comparable aux indexeurs humains. Il est hors de notre propos de présenter une synthèse ou une évaluation des différentes méthodes développées<sup>3</sup>, mais soulignons qu'à la suite d'un recensement des travaux menés sur l'indexation et la condensation automatiques, Lancaster (2003) conclue "it will probably be a very long time before machines are intelligent enough to completely replace humans in these important activities, if indeed they ever are" (p.143).

L'impact réel du travail de l'indexeur sur l'efficacité des systèmes d'information soulève la nécessité de réexaminer cette tâche afin de mieux saisir les processus intellectuels impliqués lors de l'indexation.

Ainsi, c'est dans cette perspective globale liée aux TIC que s'inscrit notre recherche sur le travail des indexeurs et plus spécifiquement sur la problématique de la cohérence interindexeurs.

## **1. Indexation de documents écrits**

Le travail de l'indexeur étant au coeur de cette recherche, il convient de présenter d'abord le contexte dans lequel prend place l'indexation afin de mettre en perspective l'importance de son rôle pour le fonctionnement des systèmes d'information. Ainsi, cette première section présente les éléments directement impliqués pour la réalisation de l'indexation. On s'attardera plus longuement à la problématique du manque de cohérence interindexeurs puisqu'elle constitue le point de départ de notre questionnement.

---

<sup>3</sup> Pour un recensement des écrits sur la question, soulignons que la deuxième édition de *Indexing and Abstracting in Theory and Practice* de Lancaster (1998) contient un chapitre sur l'indexation et la condensation automatique. Lancaster (2003) traite également des approches plus récentes.

L'indexation étant avant tout une tâche intellectuelle, la question de la représentation du contenu du document sera également abordée.

### 1.1 Chaîne documentaire et repérage de l'information

La fonction principale d'un système documentaire consiste à représenter le contenu d'un document de façon succincte et formalisée, à en conserver une trace (c'est-à-dire le produit de l'analyse) et à en permettre le repérage. Il s'agit donc d'une entreprise d'archivage et de mise à disposition de l'information qui s'articule à l'intérieur d'un vaste cadre, celui de la chaîne documentaire. La figure 1 présente plusieurs des opérations de la chaîne documentaire combinées au schéma du repérage de l'information afin d'illustrer le circuit parcouru par l'information écrite de son état original à sa forme accessible dans une base de données bibliographique informatisée.

La chaîne documentaire débute par la sélection et l'acquisition des documents, ce qui constitue le fonds documentaire. Les documents sont sélectionnés principalement en fonction d'une discipline particulière ou du sujet dont ils traitent, mais également à partir de critères, tels le type de document, la langue, la source. Dans une certaine mesure, l'étendue d'une base de données répond au rapport coût/bénéfice. Par exemple, pour la couverture de domaines spécialisés, la sélection pourra se limiter aux périodiques publiant le plus d'articles concernant le(s) domaine(s) couvert(s) par la base.

Il est à souligner que la quantité et la nature des documents acquis sont tributaires des objectifs visés par le centre documentaire. Une fois le fonds constitué, il doit être analysé et représenté dans un système formel pour faciliter le classement et le repérage.

Pour représenter les documents sélectionnés dans un système formel, chaque document subit un double traitement, descriptif et analytique. Également nommé catalogage, le traitement descriptif consiste à noter les éléments bibliographiques du document (titre, auteur, éditeur, nombre de pages, année de parution, etc.) et à permettre le repérage par ces mêmes éléments.

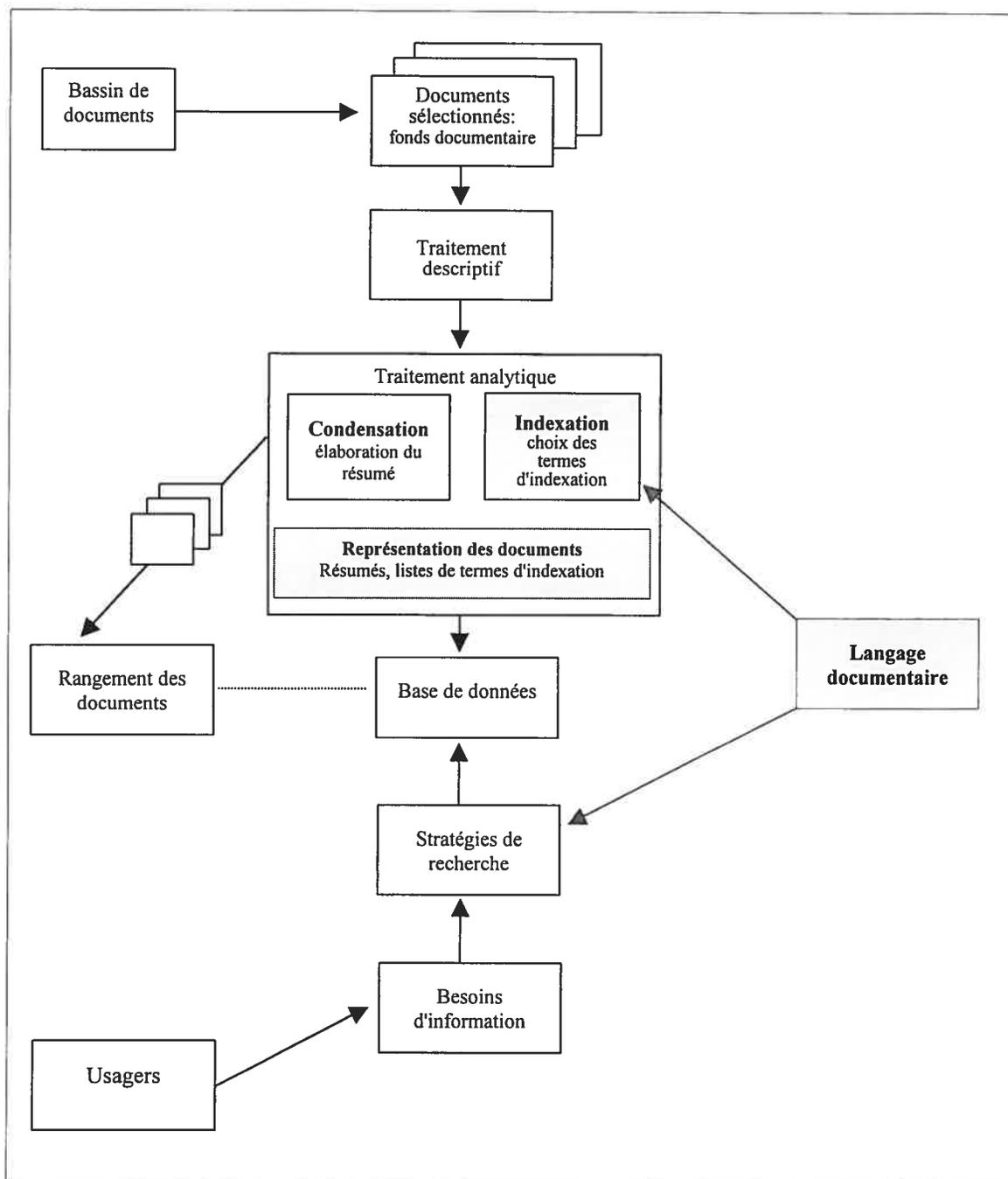


Figure 1  
Chaîne documentaire et repérage de l'information : Rôle de l'indexation.

Note. Figure adaptée et traduite de Lancaster, F. W. (1998). *Indexing and Abstracting in Theory and Practice*. Champaign, IL: Graduate School of Library and Information Science, University of Illinois (p.2).

Le traitement analytique ou analyse de contenu "vise à substituer au texte de départ un texte d'arrivée, plus court et mieux organisé, plus économique à mémoriser et à manipuler" (Bertrand-Gastaldy, 1986, p.4). Pour les fins de la recherche, nous laissons de côté la condensation (production de résumé) qui constitue une problématique en soi, pour ne nous occuper que de l'indexation. Cette analyse vise une représentation du contenu du document généralement effectuée à l'aide d'une liste de termes autorisés (mots-clés, vedettes-matière, descripteurs) issus d'un langage documentaire (répertoire vedettes-matière, thésaurus) ou du langage naturel. Spécifions que l'indexation permet une recherche d'information par sujets contrairement à une recherche à partir d'éléments bibliographiques connus, tels auteurs et titres. L'indexation étant présentée de manière exhaustive plus loin dans ce chapitre, sa présentation ne sera pas élaborée immédiatement.

Une fois l'analyse complétée, les résultats sont enregistrés dans la base de données de façon organisée sur un support ordiolingue, l'information pouvant ainsi être exploitée à des fins de recherche. Quant aux documents originaux, ils sont habituellement rangés sur les rayons du centre de documentation.

Précisons qu'il existe un lien étroit entre l'opération d'indexation et celle de la recherche documentaire puisque toutes deux s'articulent autour d'un même noyau qu'est le langage documentaire. Pour combler son besoin d'information, l'utilisateur doit élaborer une stratégie de recherche qui lui permettra de repérer les références pertinentes en réponse à sa requête.

Pour ce faire, l'utilisateur doit questionner la base avec les mêmes termes ayant servi à l'indexation pour que ce document lui soit signalé au repérage, c'est-à-dire qu'il doit choisir parmi les termes du langage documentaire, ceux qu'il croit les plus susceptibles de lui fournir des références reliées au sujet qu'il recherche et interroger la base de données avec ces termes. Le langage documentaire constitue alors le seul lien entre l'utilisateur et la représentation des documents ou liste de termes d'indexation, également appelée texte secondaire qui, à son tour, fournit l'accès aux documents originaux ou textes primaires.

## 1.2 Langages documentaires

Roy (1985) définit un langage documentaire comme "un ensemble de termes, et le cas échéant, de procédés syntaxiques conventionnels, utilisés pour représenter le contenu des documents à des fins de classement ou de recherche de ces documents" (Roy, 1985 in Bertrand, 1993, p.10). Les langages documentaires peuvent prendre plusieurs formes, les plus répandues étant les systèmes de classification, les listes d'autorité de vedettes-matière et les thésaurus.

Ces langages sont également classés selon deux aspects fondamentaux, soit la coordination et le contrôle. La notion de coordination se rapporte au fait de regrouper plusieurs termes en une seule expression servant à décrire le contenu du document ou formulée pour effectuer la recherche documentaire. Notons qu'un langage peut être précoordonné ou postcoordonné.

Un langage est précoordonné lorsqu'il coordonne avant le stockage les différents concepts qui forment un sujet: "La coordination des concepts est effectuée avant l'enregistrement et le stockage de la représentation formalisée du contenu, au moment de l'analyse, de façon à traduire en une seule expression, l'ensemble du sujet tel qu'il a été traité par l'auteur" (Bertrand-Gastaldy, 1990, p.22). Par exemple, il faudra utiliser la vedette-matière "Québec (Province) – Histoire" pour représenter un document portant sur l'Histoire du Québec. Les deux concepts élémentaires formant le sujet (Québec et Histoire) sont précoordonnés, c'est-à-dire qu'ils sont associés dans cet ordre au moment de l'indexation.

Par opposition, un langage postcoordonné permet à l'utilisateur de relier plusieurs termes au moment de la recherche : "La coordination s'effectue après l'enregistrement et le stockage, avant le repérage, au moment de la recherche documentaire" (Bertrand-Gastaldy, 1990, p.27). Ces systèmes traduisent en général non pas des classes ou des sujets mais des concepts. Ainsi, il revient à l'utilisateur d'associer les termes qui exprimeront un sujet correspondant à son besoin. Par exemple, un utilisateur cherchant des documents traitant

de *l'administration des bibliothèques* repérera les documents indexés avec les termes Québec / Bibliothèque / Administration et ceux indexés avec les termes Bibliothèque / Administration / Usager.

La notion de contrôle se traduit par l'emploi d'un vocabulaire contrôlé ou libre. Le vocabulaire contrôlé n'utilise qu'une "seule formulation acceptée pour un domaine de la connaissance, un sujet, un concept. Un réseau de relations plus ou moins élaboré regroupe les éléments voisins, les inclut dans des catégories plus vastes" (Bertrand-Gastaldy, 1990, p. 30). Ce contrôle vise entre autres, à réduire les risques d'ambiguïtés, ainsi qu'à éliminer les problèmes de polysémie, de synonymie, etc. Quant au vocabulaire libre, il ne définit aucune liste de termes autorisés *a priori*, les termes sont extraits tels qu'ils ont été trouvés dans les documents ou selon la formulation jugée appropriée par l'indexeur. Selon les caractéristiques du langage documentaire utilisé, le contenu des documents est représenté et structuré différemment, ce qui entraîne des conséquences pour le repérage.

Ainsi, un vocabulaire libre astreint l'utilisateur à penser à tous les équivalents possibles d'une clé d'accès s'il veut effectuer une recherche exhaustive, tandis qu'un langage contrôlé, tel le vocabulaire d'un thésaurus, fournit à l'utilisateur un cadre conceptuel du domaine et permet de minimiser le temps requis pour le choix des termes d'interrogation.

La Classification Décimale Universelle (CDU) est un exemple de langage précoordonné ayant un vocabulaire contrôlé. Ce système de classification utilise un code qui permet de situer le document à l'intérieur d'une discipline et de ses subdivisions (Tableau I). Généralement, pour ce type de langage (systèmes de classification encyclopédiques ou spécialisés, systèmes de vedettes-matière), "chaque indice ou vedette-matière est la seule étiquette autorisée pour représenter un sujet et, du fait de sa place dans une hiérarchie ou de sa formulation, réfère à un sujet valable dans un domaine plus qu'à un concept, ..." (Courrier, 1976, p.185).

Ce type de langage est facile d'utilisation autant pour le documentaliste que pour l'utilisateur et la numérotation des hiérarchies permet d'élargir ou de spécifier à volonté une classe donnée. Développé au XIX<sup>e</sup> siècle, le classement hiérarchique de la CDU a été

élaboré pour le traitement de fonds encyclopédiques et sa principale utilité est de permettre le classement des livres sur les rayons des bibliothèques. Il ne permet pas une représentation détaillée des contenus et se limite au sujet général.

Tableau I. Extrait de la Classification Décimale Universelle (CDU).

Les dix classes de base de la CDU.	
0	Généralités. Documentation. Bibliographie. Bibliothéconomie.
1	Philosophie. Éthique. Psychologie.
2	Religion. Théologie.
3	Sciences sociales. Droit.
4	(classe vacante)
5	Mathématiques.
6	Sciences appliquées. Médecine. Technologie.
7	Arts. Loisirs. Sport.
8	Langues. Linguistique. Littérature.
9	Géographie. Biographie. Histoire.
Les divisions et subdivisions d'une classe, du plus général au plus spécifique.	
6	Technologie.
62	Engineering. Technologie en général.
621	Engineering mécanique. Technologie nucléaire. Engineering électrique.
621.3	Engineering électrique.
621.39	Télécommunication. Télégraphie. Radio-communications. Facsimilé. Télévision. Télécommande.
621.396	Radio-communication. Transmission du son.

*Note.* Cet extrait de la CDU provient de Chaumier, J. (1988). *Le traitement linguistique de l'information*. Paris: Entreprise moderne d'édition (p.60).

Ce système ne permet que des approximations de représentations, créant un *bruit*<sup>4</sup> élevé au repérage (Swan Hill, 1984 in Bertrand-Gastaldy, 1990). De plus, cette structure monolithique de l'ensemble des connaissances ne permet plus de rendre compte adéquatement de la complexité des documents produits à notre époque où les approches multidisciplinaires abondent (Bureau canadien des archivistes, 1992). Mais une analyse peu profonde, effectuée rapidement et qui ne requiert de l'indexeur qu'un minimum d'expérience et de connaissance du domaine entraîne des coûts relativement bas.

<sup>4</sup> Les notices repérées qui sont non pertinentes par rapport à la question constituent le bruit.

Devenu l'outil privilégié des systèmes documentaires au cours des années '70, le thésaurus est généralement constitué d'un vocabulaire contrôlé, postcoordonné et regroupe un ensemble d'éléments, tels descripteurs, termes équivalents et diverses relations qui précisent l'environnement sémantique des termes (Chaumier, 1988). Ce langage documentaire s'applique à des domaines spécialisés, la structure assure une meilleure définition de chaque terme et permet des recherches à différents degrés de généralité ou de spécificité. C'est en partie pour répondre à certaines des lacunes des classifications que les thésaurus ont été développés. En effet, un besoin croissant pour une indexation adéquate des documents spécialisés a incité les *scientifiques de l'information* à mettre au point un langage permettant une représentation plus exhaustive et précise du contenu des documents (Bureau canadien des archivistes, 1992; Chaumier, 1988). Les descripteurs d'un thésaurus peuvent être moins spécifiques que les termes ou expressions d'un langage précoordonné, mais la postcoordination de plusieurs descripteurs rend la représentation des documents et de la question plus précise. En conséquence, l'utilisation de thésaurus apparaît comme une solution aux problèmes de *bruit* et de *silence*<sup>5</sup> lors du repérage. Cependant l'adoption de ce langage entraîne des coûts élevés puisque la main-d'oeuvre doit être qualifiée pour l'élaboration et la mise à jour du thésaurus, et les indexeurs doivent avoir une connaissance appropriée du domaine couvert.

Contrairement aux classifications traditionnelles, un thésaurus introduit entre les termes diverses relations: relation d'équivalence, relation hiérarchique, relation d'association, relation de définition. Les exemples suivants sont extraits du thésaurus Envirodoq (Centre de documentation et de renseignements du MENVIQ, 1993)<sup>6</sup>, soit l'un des deux thésaurus utilisés pour la présente recherche.

---

<sup>5</sup> Les notices pertinentes existant dans le fonds documentaire, mais non signalées au repérage constituent le silence.

<sup>6</sup> Afin d'alléger le texte et parce que le thésaurus est l'outil de travail, le nom d'auteur et l'année de parution seront omis dorénavant. Le thésaurus Envirodoq est développé et utilisé par le Ministère de la Faune et de l'Environnement du Québec. Comme ce thésaurus est l'un des deux thésaurus servant pour l'expérimentation de la présente recherche, il nous a semblé pertinent de présenter immédiatement au lecteur certaines de ses caractéristiques, qui d'ailleurs sont partagées par la plupart des thésaurus.

La relation d'équivalence indique le renvoi vers le descripteur des termes synonymes ou quasi-synonymes qui ne peuvent être utilisés pour l'indexation. La forme inverse est également présente :

Usine épuration EM (employer) Station épuration	Station épuration EP (employer pour) Usine épuration
--	---

La relation hiérarchique permet de situer les termes en fonction des rapports d'emboîtements qui les unissent, soit un rapport de superordination (terme générique) ou de subordination (terme spécifique):

Réchauffement terre TG (terme générique) Changement climatique TS (terme spécifique) Effet serre	Effet serre TG Gaz carbonique Rayonnement solaire Réchauffement terre
--	--

La relation associative soulève les analogies entre deux ou plusieurs termes de familles apparentées. Ce type de relation est souvent indiqué par l'expression *voir aussi (VA)*. Comme pour les relations précédentes, la forme inverse est présente :

Pluie TA (terme associé) Averse Crue Érosion Fréquence pluie Pluviomètre	Averse TA Crue Érosion Pluie Orage
---	---

On retrouve également des notes d'application (NA), dites relations de définition qui servent à préciser le sens d'un descripteur et son contexte d'utilisation afin d'éviter les problèmes de synonymie :

Station épuration NA Pour traiter les eaux usées	Station purification NA Pour rendre l'eau potable
---	--

Il existe plusieurs représentations pour le contenu des thésaurus (les plus fréquemment rencontrées étant les présentations alphabétique, alphabétique permutée, thématique, hiérarchique et graphique) regroupées en deux grandes catégories: les thésaurus

dictionnaires ou thésaurus à structure interne, et les thésaurus à représentation graphique ou thésaurus à structure externe. Généralement, plusieurs représentations coexistent au sein d'un même thésaurus constituant des listes complémentaires. Par exemple, le thésaurus Infoterra (Programme des Nations Unies pour l'Environnement, 1990)<sup>7</sup> relatif au domaine de l'environnement comporte quatre listes :

1) *Le sommaire des catégories et des sous-catégories* donne une vue d'ensemble du contenu du thésaurus, mais ne contient aucun terme particulier:

Code	10000	ATMOSPHERE	<u>Catégorie</u>
	10100	Composition de l'atmosphère	<u>Sous-catégorie</u>
	10200	Phénomènes atmosphériques	
	10300	Pollution atmosphérique	
	10400	Climat et changement climatique	

2) *La liste des termes d'après leur numéro de code* présente chaque terme une seule fois dans une catégorie ou une sous-catégorie selon son appartenance principale. Cette liste sert à chercher un terme lorsque seul le numéro de code est connu et à indiquer des groupes de termes appartenant à une même sous-catégorie:

Code	10000	ATMOSPHERE	<u>Catégorie</u>
	10100	Composition de l'atmosphère	<u>Sous-catégorie</u>
	10110	Qualité de l'air	<u>Terme</u>
	10120	Chimie de composition de l'atmosphère	
	10130	Particules atmosphériques	
	10140	Dioxyde de carbone	

3) *La liste de termes d'après leur catégorie* énumère tous les termes du thésaurus en ordre alphabétique par catégorie et sous-catégorie avec les relations génériques, spécifiques, associatives et équivalentes. Cette liste sert à déterminer et sélectionner les termes les plus pertinents dans la catégorie ou sous-catégorie recherchée:

<sup>7</sup> Afin d'alléger le texte et parce que le thésaurus est l'outil de travail, le nom d'auteur et l'année de parution seront omis dorénavant. Le thésaurus Infoterra a également été utilisé pour notre expérimentation. Sa structure est décrite ici pour les raisons évoquées précédemment.

10000	<b>ATMOSPHERE</b>	<u>Catégorie</u>
10300	<b>Pollution atmosphérique</b>	<u>Sous-catégorie</u>
	<b>Chlorofluorocarbones (10340)</b>	<u>Terme</u>
Employé pour (EP):	CFC's Fréons	
Terme générique (TG):	Composés organo-halogènes	
Terme associés (TA):	Aérosols Couche d'ozone	

4) *La liste alphabétique des termes* présente chaque terme par ordre alphabétique avec toute la composition relationnelle et les formes inversées des descripteurs à concepts multiples. Cette liste permet de vérifier un terme particulier:

Terme	<b>28148</b>	<b>Cultures hydroponiques</b>
Note d'application		<i>NA: Culture dans l'eau sans sol</i>
Terme	<b>17139</b>	<b>Cultures selon les courbes de niveau</b>
Employé pour		EP: Cultures en terrasses
Terme	<b>10240</b>	<b>Cycle du carbone</b>
Terme proscrit		Cyclisme
Employé		EM: Bicyclettes

Chaque langage documentaire possède des avantages et des inconvénients, et malgré leurs qualités, les thésaurus n'échappent pas à la critique. "Le caractère fini de la liste qu'est avant tout un thésaurus donne à penser que les valeurs sémantiques de chacun de ses éléments se délimitent les unes par rapport aux autres" (Dachelet, 1990, in Bertrand, 1993, p.16). En d'autres termes, la liste de descripteurs attribués à un document permet de le situer seulement à l'intérieur d'une représentation finie du domaine auquel il appartient. De plus, cette liste n'a de valeur qu'en fonction de l'ensemble des listes enregistrées dans la base de données bibliographiques.

Comme le souligne Bertrand-Gastaldy (1986), il n'est pas facile pour un centre de documentation de choisir un langage documentaire<sup>8</sup> :

Le choix réside en effet de la prise en compte de nombreux éléments en interrelation à l'intérieur de systèmes et de sous-systèmes complexes et la décision n'est bien souvent que le résultat d'un compromis entre tendances diverses, voire contradictoires. (p.3)

<sup>8</sup>Compte tenu des objectifs de cette recherche, nous nous limitons à la présentation des éléments nécessaires et suffisants à la compréhension de la recherche. Une revue exhaustive des divers thésaurus, leur élaboration, leur évaluation et leur gestion informatique est présentée dans Chaumier (1988).

Quel que soit le langage documentaire utilisé, le lien entre l'utilisateur et les documents originaux s'établit par le texte secondaire, c'est-à-dire le résultat du travail de l'indexeur. Il est donc pertinent de se tourner vers l'opération clé des systèmes d'information, soit l'indexation.

### 1.3 Description de la tâche d'indexation

La norme internationale ISO (Organisation internationale de normalisation [ISO], 1985)<sup>9</sup> définit l'indexation en ces termes :

On extrait les notions des documents par un processus d'analyse intellectuelle, puis on les transcrit à l'aide d'éléments du langage d'indexation. L'analyse et la transcription doivent toutes deux être faites avec l'aide d'outils d'indexation tels que thésaurus et plans de classement. (p.2)

Du point de vue de sa finalité, l'indexation est destinée à permettre une recherche efficace des informations dans un fonds documentaire.

L'indexation comprend essentiellement trois étapes qui peuvent se chevaucher dans la pratique (ISO, 1985):

- examen du document et définition de son contenu,
- identification et sélection des notions principales du contenu,
- choix des termes d'indexation, ou traduction en langage documentaire (p.2)

Les écrits portant sur l'indexation segmentent habituellement le processus d'indexation en deux étapes, soit: la reconnaissance des notions principales représentant le contenu essentiel du document et la représentation de ces concepts en termes issus du langage documentaire.

Étant donné les contraintes de production, une lecture complète du document est rarement praticable. Par contre, l'indexeur doit s'assurer de ne négliger aucune information.

---

<sup>9</sup> Les normes reconnues comme les plus importantes pour la description des méthodes pour l'indexation des documents sont les normes ISO 5963-1985 de l'Organisation internationale de normalisation (ISO, 1985), BS 6529: 1984 de la British Standards Institution (BSI, 1984) et NF Z 47-102 de l'Association française de normalisation (AFNOR, 1978). Nous nous référons à la norme ISO (1985) puisque les deux autres normes s'en inspirent directement.

Il examine alors le document en le feuilletant rapidement et en consultant les sections les plus informatives : titre, résumé, table des matières, introduction, conclusion, début de chapitres et de paragraphes, tableaux, figures, et les mots mis en valeur par une typographie particulière. Une fois le relevé de tous les éléments complété, l'indexeur doit identifier les notions essentielles et représentatives du contenu. Cependant, "l'indexeur ne doit pas nécessairement retenir, comme termes d'indexation, tous les concepts recensés lors de l'examen du document" (ISO, 1985, p.3). La finesse de la sélection dépend principalement de la nature des documents, de l'étendue du langage documentaire, de la valeur potentielle pour les utilisateurs et des politiques particulières d'un centre de documentation. La norme ISO émet quelques recommandations à cet égard :

- La couverture de l'indexation ne doit pas être limitée de façon trop étroite. Avec le développement des réseaux d'information, ... il est recommandé aux indexeurs de littérature scientifique et technique de ne pas négliger les autres facettes d'un sujet, par exemple son aspect social ou économique.
- Le critère principal dans la sélection des notions, doit toujours être la valeur potentielle d'une notion en tant qu'élément dans l'expression du contenu d'un document et dans sa recherche documentaire.
- Le nombre de termes ou de descripteurs ... ne doit pas être limité arbitrairement, ... mais déterminé entièrement par la quantité d'information contenue dans le document en liaison avec les besoins supposés des utilisateurs.
- Les notions doivent être identifiées d'une façon aussi spécifique que possible. (p.3)

Pour favoriser une sélection et une traduction appropriées, la norme ISO souligne également que :

- L'indexeur doit avoir une bonne connaissance du domaine couvert par les documents.
- Il doit être familiarisé avec les outils d'indexation ... et être conscient du fait que ces outils peuvent imposer certaines contraintes. (p.4)

Finalement, l'indexeur doit traduire les notions essentielles en termes autorisés d'un langage documentaire. Pour ce faire, il consulte le langage sous le terme choisi afin d'en vérifier la validité. Il décide alors de conserver le terme ou continue la consultation en examinant le contexte du terme consulté (TG, TS, TA, et renvois dans le cas d'un thésaurus) pour évaluer le terme le plus approprié pour rendre compte de la notion retenue. Il arrive qu'un terme jugé essentiel par l'indexeur ne se trouve pas dans le langage. Dans ce cas,

l'indexeur peut proposer un nouveau terme (appelé candidat descripteur pour les thésaurus) qu'il inscrit dans un champ particulier, lequel terme sera évalué et possiblement intégré au vocabulaire lors de la mise à jour du langage. La liste de termes officiels complétée est intégrée à la base de données et servira au repérage par sujets (descripteurs, vedettes-matière, etc.).

L'indexation est avant tout une tâche intellectuelle et de surcroît le maillon le plus coûteux de la chaîne documentaire, le temps consacré à l'indexation d'un document pouvant varier d'une durée de quelques minutes à parfois plus d'une heure. Le succès d'une recherche documentaire et la qualité du système documentaire dépendent en majeure partie de la justesse et de la richesse de la représentation des documents :

En effet, l'indexation conditionne la valeur d'un système documentaire. Une mauvaise indexation ou une indexation insuffisante représente 90 % des causes essentielles de l'apparition, lors d'une recherche, de bruits ou de silences. Les 10 % résiduels seront dus à des causes mécaniques. (Chaumier, 1988, p.54.)

Il devient alors primordial de s'assurer de la fiabilité de l'indexation qui peut être évaluée en fonction de plusieurs critères de qualité.

#### 1.4 Critères de qualité

La qualité de l'indexation peut être évaluée en fonction des taux de rappel et de précision (associés aux notions de bruit et de silence) à la suite du repérage. Le taux de rappel représente le pourcentage de documents pertinents retrouvés par rapport au nombre total de documents pertinents contenus dans le système, et le taux de précision se rapporte au pourcentage de documents pertinents repérés par rapport au nombre total de documents retrouvés au repérage.

L'indexeur doit cependant être en mesure de juger par lui-même la qualité de son indexation dès la phase d'analyse complétée, c'est-à-dire avant le stockage et le feed-back des résultats de la recherche documentaire. C'est pourquoi il devra porter attention aux critères de qualité suivants :

- Exhaustivité (ou profondeur)<sup>10</sup>

L'exhaustivité est liée à une mesure quantitative et correspond au nombre de concepts retenus et traduits pour caractériser le contenu intégral d'un document. La norme ISO recommande "d'identifier toutes les notions qui ont une valeur potentielle pour les utilisateurs d'un système d'information dans un document" (ISO, 1985, p.3).

- Précision (ou spécificité)

La précision correspond au degré de généralité ou de spécificité des termes d'indexation choisis par l'indexeur. "La spécificité de l'indexation n'implique pas la recherche des descripteurs les plus spécifiques mais l'adéquation entre la spécificité du concept et celle du descripteur" (Bonville, 1983 in Patch-Neveu, 1989, p.26). Ce critère est étroitement lié à l'exactitude puisqu'il correspond à la capacité d'un terme à discriminer un petit nombre de documents, c'est-à-dire de réduire aux seuls documents très pertinents les documents repérés.

- Exactitude (ou véracité)

L'exactitude correspond à la fidélité avec laquelle le contenu d'un document est traduit en termes autorisés. L'indexeur doit s'efforcer "de ne pas trahir le contenu; il ne doit pas signaler des notions et des relations entre notions qui ne sont pas présentes dans le fonds, la série, le dossier ou la pièce" (Bureau canadien des archivistes, 1992, p. 69). La confiance des usagers dans le système documentaire dépend en grande partie du respect de l'exactitude dans l'indexation.

Évidemment, tous ces critères dépendent en grande partie de la qualité du langage documentaire. Par exemple, les critères d'exhaustivité et d'exactitude dépendent des possibilités offertes par l'étendue et la spécificité même du vocabulaire du langage, mais dépendent également de la représentation que l'indexeur se fait du document.

Bien que le respect de ces critères devrait permettre à l'indexeur de produire une indexation de qualité, ils sont difficiles à mesurer et surtout à opérationnaliser (Rolling,

---

<sup>10</sup> Les termes entre parenthèses sont considérés comme équivalents dans la littérature et sont utilisés de manière interchangeable dans le présent exposé.

1981). En fait, il n'existe aucun critère permettant de juger de la validité absolue d'une indexation, d'une part, parce que la notion de *contenu d'un document* est relativement indéfinissable et d'autre part, parce qu'il n'existe pas de réponse fixe à la représentation d'un document.

Cependant, il est possible de vérifier indirectement la qualité relative d'une indexation en observant s'il existe une certaine fidélité dans le processus d'indexation, c'est-à-dire jusqu'à quel point deux personnes produiront des listes similaires pour un même document. Ceci renvoie au critère de la cohérence, plus facilement observable, mesurable et ayant fait l'objet de nombreuses études.

### 1.5 Problématique de la cohérence interindexeurs

On entend par cohérence interindexeurs "la capacité chez plusieurs indexeurs d'analyser un même document de la même façon et de représenter les mêmes concepts par les mêmes termes" (Bertrand-Gastaldy, 1986, p.11). Le taux de cohérence interindexeurs est une façon indirecte de savoir jusqu'à quel point l'indexation peut être fiable.

Le calcul du taux de cohérence interindexeurs est le rapport, exprimé en pourcentage, entre le nombre de termes communs de deux indexeurs, multiplié par deux, sur le nombre total de termes choisis par ces mêmes indexeurs<sup>11</sup>. Conséquemment, le degré d'accord varie de 0 à 1: 0 s'il n'y a aucun recouvrement entre les termes sélectionnés par les deux indexeurs, et 1 s'ils ont retenus exactement les mêmes termes.

Si plusieurs indexeurs d'expertise égale arrivent au même résultat, on ne peut conclure directement à la validité de leurs indexats, mais cet accord fournit un indice probant sur la qualité relative de l'indexation. Cependant, si les indexeurs produisent des listes de termes différents, on questionnera alors la valeur respective de chacune des

<sup>11</sup> La formule utilisée est la suivante :  $2C / (A + B)$ .

A = nombre total de termes de l'indexeur A.

B = nombre total de termes de l'indexeur B.

C = nombre de termes communs aux indexeurs A et B.

Précisons que les calculs de la cohérence interindexeurs sont exposés en détail au chapitre suivant.

représentations. Le degré de cohérence interindexeurs permet de déterminer dans une certaine mesure, le degré de confiance que l'utilisateur pourra accorder au système documentaire par rapport aux résultats de sa recherche.

Bien que concrètement un système documentaire ne contienne pas plusieurs indexats pour le même document, les études sur la cohérence démontrent que des documents traitant de sujets similaires pourront être indexés avec des ensembles de termes si différents que leur ressemblance ne pourra être perçue par la comparaison des listes de termes d'indexation (Zunde et Dexter, 1969).

Comme il a été mentionné précédemment, l'indexation est régie par des normes et des critères auxquels doivent se conformer les indexeurs. Cependant, il existe une inéquation entre les termes choisis par différents indexeurs pour représenter un même document, résultant en un repérage souvent inadéquat et surtout incomplet, ce qui crée une incertitude constante pour l'usager quant à l'efficacité du repérage. Comme le précise Rolling (1981):

If one compares indexing terms assigned by any two indexers to the same documents, one discovers that, as a rule, the indexers differ considerably in their judgment as to which terms reflect the contents of the document most adequately. (Rolling, 1981, p.69).

Depuis les années '50, les chercheurs en sciences de l'information (SI) ont tenté d'éclaircir la problématique du manque de cohérence interindexeurs. La recension de plus de 35 études (Zunde et Dexter, 1969; Leonard, 1977; Markey, 1984) révèle un nombre important de facteurs qui influencent la cohérence et démontre que le degré de cohérence demeure relativement faible. Soulignons que Markey (1984) rapporte les résultats de 25 études dont le tiers seulement obtient un taux de cohérence supérieur à 50 %.

Zunde et Dexter (1969) ont identifié une liste de vingt-cinq facteurs susceptibles de faire varier la cohérence interindexeurs, lesquels sont répertoriés en trois catégories, soit les variables sémantiques, pragmatiques et environnementales.

Il semble normal que les variables sémantiques constituent le groupe ayant fait l'objet du plus grand nombre d'études puisqu'elles sont reliées directement au contenu des

documents, aux outils d'indexation et aux contraintes pour le choix des termes. Selon Zunde et Dexter (1969), ce type de variables déterminent les variations d'interprétation de la signification des expressions, des énoncés, des phrases et des mots du document indexé ainsi que la signification des termes d'indexation lui étant assignés.

Tableau II. Facteurs potentiels de variation de la cohérence interindexeurs.

Variables sémantiques	Variables pragmatiques	Variables environnementales
1- Vocabulaire de l'auteur 2- Lisibilité du texte * 3- Domaine du sujet * 4- Spécificité du discours 5- Style et clarté de la présentation 6- Redondance de l'information dans le texte 7- Méthode d'indexation 8- Contraintes imposées sur le choix des termes d'indexation 9- Outils d'indexation de nature sémantique tels: thésaurus, répertoires vedettes-matière *	10- Scolarité des indexeurs * 11- Maturité, âge 12- Personnalité 13- Acquis personnels 14- Aptitude à l'indexation 15- Motivation au travail 16- Disposition (humeur, caractère...) 17- Fatigue, santé, etc. 18- Contraintes psychologiques 19- Importance du jugement relatif à la structure du document	20- Bruit dans la zone de travail 21- Ambiance thermique et climatique * 22- Éclairage 23- Interruptions dans le travail 24- Contraintes de temps 25- Genre d'équipement technique (support, etc.)

Note. D'après Zunde, P., Dexter, M.E. (1969). Factors Affecting Indexing Performance, *Proceedings of the 32nd Annual Meeting of the American Society for Information Science* (pp. 313-322). Westport, CT: Greenwood.

\* Variable ayant fait l'objet d'une ou de plusieurs recherches.

Quant aux variables pragmatiques, elles influencent principalement l'importance des jugements et les jugements de valeur dans l'indexation. L'influence de ces facteurs semble se manifester dans les préférences individuelles pour le choix de certains termes et le nombre de termes assignés, supposant qu'il n'y a pas de différence d'interprétation du sens des termes (Zunde et Dexter, 1969).

Toujours, selon Zunde et Dexter (1969), un examen exhaustif de l'influence de ces variables permettrait de mettre au point un test d'aptitude à l'indexation qui permettrait d'identifier les individus les plus aptes à devenir de bons indexeurs. Cependant, plusieurs notions, telles la motivation, l'aptitude ou la personnalité devraient être définies de manière beaucoup plus précise par les auteurs (Bertrand, 1993).

Tableau III. Principaux résultats des études sur les facteurs affectant la cohérence interindexeurs.

HYPOTHÈSES	RÉSULTATS		
	Hypothèse confirmée	Résultats opposés	Aucune différence
<b>Expertise de l'indexeur</b> Les indexeurs expérimentés possèdent un niveau de cohérence supérieur à celui des indexeurs inexpérimentés.	<ul style="list-style-type: none"> <li>• Slamecka et Jacoby (1962)</li> <li>• Gehring (1964)</li> <li>• Case Western Reserve University (1968)</li> <li>• Leonard (1975) 2 études</li> </ul>		<ul style="list-style-type: none"> <li>• Fried et Prevel (1966)</li> </ul>
<b>Utilisation d'un vocabulaire contrôlé</b> La cohérence interindexeurs augmente lorsque les termes sont choisis au sein d'un vocabulaire contrôlé.	<ul style="list-style-type: none"> <li>• Slamecka et Jacoby (1963) 2 études</li> <li>• Korotkin et Oliver (1964)</li> <li>• Fried et Prevel (1966)</li> <li>• Herr (1970)</li> <li>• Tarr et Broko (1974)</li> </ul>	<ul style="list-style-type: none"> <li>• Slamecka et Jacoby (1963)</li> </ul>	
<b>Expertise dans l'utilisation du langage</b> Les indexeurs familiers avec le langage documentaire utilisé obtiennent une cohérence plus élevée que les indexeurs non familiers.	<ul style="list-style-type: none"> <li>• Bates (1977)</li> </ul>		
<b>Connaissance du domaine</b> Les spécialistes d'un domaine ont une cohérence supérieure aux non spécialistes.	<ul style="list-style-type: none"> <li>• Korotkin et Oliver (1964)*</li> <li>• Leonard (1975)*</li> </ul>	<ul style="list-style-type: none"> <li>• Korotkin et Oliver (1964)*</li> </ul>	<ul style="list-style-type: none"> <li>• Leonard (1975)</li> <li>• Bates (1977)</li> </ul>
<b>Profondeur de l'indexation</b> La cohérence est plus forte lorsque le nombre de termes assignés diminue.	<ul style="list-style-type: none"> <li>• Slamecka et Jacoby (1962)</li> </ul>		<ul style="list-style-type: none"> <li>• Harris, Rayward et Svenonius (1966)</li> </ul>

Note. Tableau compilé à partir de Markey, K. (1984). Inter-indexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials. *Library and Information Sciences Research*, 6, 155-177.

\*Indique des résultats non significatifs.

Selon Zunde et Dexter (1969), les variables environnementales ne concernent pas l'indexeur ou les documents indexés mais peuvent influencer la réalisation de l'indexation, bien qu'une étude sur l'influence des conditions climatiques n'a indiqué aucun effet significatif.

D'autres variables non répertoriées par Zunde et Dexter (1969) ont été mises en relief dans les recensions d'écrits faites par Leonard (1977) et Markey (1984), notamment

l'expérience de l'indexeur, la connaissance du domaine<sup>12</sup> des documents à indexer, la connaissance des outils d'indexation, le type de langage documentaire utilisé et la profondeur de l'indexation. Sans présenter les résultats de ces études de manière exhaustive, ils sont regroupés dans le Tableau III selon qu'ils confirment l'hypothèse, qu'ils y sont opposés, ou qu'ils ne révèlent aucune influence de la variable testée sur la cohérence interindexeurs. Étant donné le nombre élevé d'études recensées par Leonard (1977) et Markey (1984), nous ne rapportons ici que les résultats les plus pertinents par rapport aux objectifs de la présente recherche.

Concernant la performance d'indexeurs novices et experts, trois des cinq études ayant révélé une plus forte cohérence chez les experts ont démontré que la cohérence n'était que légèrement plus élevée chez ces derniers. Toutefois, l'étude de Fried et Prevel (1966, in Leonard, 1977) ne démontre aucune différence significative entre experts et novices.

La majorité des études portant sur l'utilisation d'un vocabulaire contrôlé indique que le taux de cohérence interindexeurs augmente avec l'emploi d'un vocabulaire contrôlé comparativement à l'emploi d'un vocabulaire libre. De façon similaire, Slamecka et Jacoby (1963, in Leonard, 1977) ont constaté que les taux obtenus croissaient lorsque des indexeurs experts utilisaient un vocabulaire contrôlé, sans toutefois pouvoir attribuer la cause à l'un ou l'autre des deux facteurs mis en relation. Cependant, très peu d'études ont exploré l'influence de l'expertise dans l'utilisation du langage documentaire. L'étude de Bates (1977, in Leonard, 1977) semble toutefois indiquer une augmentation du taux de cohérence lorsque l'outil est connu des indexeurs.

Quant aux études s'étant attardées à l'influence de la connaissance du domaine, les résultats démontrent qu'il n'existe aucun effet apparent de la connaissance du domaine sur la cohérence interindexeurs. Ajoutons que MacMillan et Welt (1961, in Leonard, 1977) rapportent que l'indexeur spécialiste sera plus enclin à choisir des termes très spécifiques de son domaine, l'expression de nuances pouvant alors mener à une chute du taux de cohérence.

---

<sup>12</sup> Chez Leonard (1977), la notion de connaissance du domaine renvoie à la formation antérieure de l'indexeur (indexer educational background).

Notons que sur quinze études ayant testé des indexeurs experts, six ont testé des indexeurs à la fois experts et spécialistes du domaine (Markey, 1984). Le peu d'études se rapportant aux experts-spécialistes nous permet de supposer que les indexeurs acquièrent probablement une connaissance du domaine des documents à indexer au fur et à mesure de leur expérience, plutôt que d'en posséder une connaissance *a priori*. Cette tendance semble être une pratique courante dans les centres de documentation, où les nouveaux indexeurs se voient attitrés à l'indexation des documents d'un domaine couvert exhaustivement par le centre. Ils prendront en moyenne de six mois à un maximum d'une année pour devenir "experts" aux yeux de leurs collègues<sup>13</sup>.

Quant à l'influence de la profondeur de l'indexation, Slamecka et Jacoby (1962, in Leonard, 1977) ont démontré que les indexeurs expérimentés attribuent significativement moins de termes que les novices et sont parallèlement plus cohérents (16 % et 12 % respectivement). Il est cependant difficile d'identifier l'influence prédominante de l'une ou l'autre des variables testées. L'augmentation de cohérence de 4 % dans l'étude de Slamecka et Jacoby (1962, in Leonard, 1977) serait-elle attribuable à l'expertise des indexeurs plutôt qu'à la profondeur de l'indexation? L'étude de Harris et al. (1966, in Leonard, 1977) n'a trouvé aucun effet du nombre de termes assignés sur la cohérence interindexeurs.

En résumé, il ressort de ces recherches que l'expertise de l'indexeur et l'utilisation d'un vocabulaire contrôlé permettent d'augmenter la cohérence interindexeurs, tandis que la connaissance du domaine et la profondeur de l'indexation ne semblent pas avoir d'effets marquants.

Leonard (1977) et Markey (1984) soulignent que tous ces résultats doivent être considérés séparément et toute comparaison doit être établie avec prudence. Les divergences considérables quant à la méthodologie employée, le mode de calcul, le traitement des variables, le type de documents, le contrôle imposé au processus d'indexation, etc., révèlent un problème de constance méthodologique rendant toute

---

<sup>13</sup> Entrevues personnelles effectuées avec les responsables de l'indexation de plusieurs centres de documentation.

généralisation difficile. Par exemple, les études évaluant le rôle d'un vocabulaire contrôlé sur la cohérence utilisaient tantôt des indexeurs experts connaissant le domaine, tantôt des indexeurs novices sans connaissance particulière du domaine. Parallèlement, les études comparant la cohérence entre des groupes d'experts et de novices se servaient de différents langages documentaires, et parfois faisaient varier les portions de texte soumises à l'indexation (titre, résumé, texte complet, etc.). Lorsque plusieurs variables sont évaluées simultanément, il devient alors difficile d'analyser comment chaque variable influence la cohérence. Il n'est donc pas surprenant que certains de ces résultats apparaissent contradictoires (Rabasse, 1989).

L'insuffisance de ces études provient d'une part de l'approche purement expérimentale choisie pour examiner la problématique de la cohérence interindexeurs. En effet, l'ensemble de ces recherches fondent leurs analyses sur le résultat final de l'indexation, la liste des termes retenus, et tentent d'attribuer à la variable testée la cause du taux plus ou moins élevé du recouvrement des listes produites par différents indexeurs pour un même document. Elles ne fournissent aucune explication du lien entre les variables testées et la variation des listes produites. D'autre part, ces études ne considèrent jamais la production de la liste elle-même. En d'autres termes, ces études ne s'attardent pas aux processus intellectuels sous-jacents à la production de ces listes.

Ainsi, les études sur la cohérence interindexeurs semblent fournir certaines pistes pour expliquer la cohérence interindexeurs peu élevée, mais ne donnent pas d'explication satisfaisante et l'on déplore toujours la situation (Saracevic, 1991). Comme le souligne Lancaster (1998), "while many consistency studies have been performed over the years, there has been little research on *why* different indexers select different terms, admittedly a more difficult type of investigation" (p.74).

Effectivement, ces vingt dernières années fournissent très peu de références concernant la cohérence interindexeurs. Quelques études plus récentes s'intéressant à la cohérence interindexeurs (Tonta, 1991; Sievert et Andrews, 1991; Bertrand, 1993) confirment le constat d'un manque de cohérence.

Il apparaît donc important de s'attarder au processus d'indexation et non juste au produit afin de mieux comprendre le problème de la cohérence interindexeurs. Le but explicite de l'indexation est de représenter le sujet ou propos d'un document et de le transposer en des termes qui en favorisent le repérage. Mais qu'est-ce que le sujet d'un document? Certains auteurs supposent l'impossibilité d'atteindre une forte cohérence interindexeurs étant donné la nature indéfinissable du sujet ou propos d'un document. En ce sens, nous tenterons d'abord de définir ce qu'est la représentation du sujet, pour ensuite nous attarder aux processus intellectuels menant à cette représentation.

### 1.6 La nature et la représentation du sujet

Sans présenter la problématique de manière exhaustive, on doit noter que la littérature en SI ne contient pas une définition claire et univoque de la nature du *propos* (*aboutness*) d'un document, notion fondamentale pour le travail de l'indexeur. Le débat oppose la permanence à l'indétermination de la nature du *propos*.

En bref, Fairthorne (1969, in Beghtol, 1986) suggère deux types de *propos*: le *propos* extensionnel (*extensional aboutness*) en tant que sujet inhérent au document, et le *propos* intentionnel (*intensional aboutness*) reflétant l'intention pour laquelle le centre de documentation a acquis le document ou la raison pour laquelle un usager en fait la demande. Boyce (1982, in Beghtol, 1986) établit une distinction similaire en parlant de sujet (*topicality*) comme étant *l'aboutness* déterminé par les sujets évidents des documents, et parle d'informativité (*informativeness*) pour exprimer le sens subjectif attribué par l'utilisateur.

Par contre, Wilson (1985, in Todd, 1992) rejette cette supposition en soutenant que la nature du *propos* est essentiellement indéterminée et que son analyse (entre autres par l'indexation) est un processus foncièrement subjectif, le point de vue d'un indexeur sur un *propos* donné étant façonné en fonction de ce processus. Maron (1977, in Todd, 1992) reconnaît également que *l'aboutness* peut être interprété de plusieurs façons et que chacun de ces points de vue amène une interprétation différente de la signification du *propos*. De

façon similaire, Oliver (1989, in Todd, 1992) argumente qu'un document peut avoir différentes significations pour le même lecteur ou indexeur à différents moments, et qu'il possède un nombre illimité de significations en fonction des perceptions reliées aux diverses utilisations qui en sont faites.

Le *propos* prend donc une signification différente selon qu'un document est considéré comme une unité isolée et finie dont le *propos* est circonscrit par cette seule unité et peut être alors analysé (ou indexé) comme tel, ou selon que le *propos* du document est considéré dans son rapport avec un ensemble de dimensions extratextuelles (but de la lecture, besoin précis d'information du lecteur, rapport avec l'ensemble des documents de la base, etc.).

Beghtol (1986) clarifie la terminologie en distinguant *propos* (*aboutness*) et sens (*meaning*) d'un document à partir des définitions proposées par Fairthorne. Cette distinction est basée sur le postulat qu'un document possède un sujet intrinsèque, son *aboutness*, qui est en partie indépendant de toute utilisation temporaire pour lequel le lecteur lui attribuerait une ou plusieurs significations:

The subject of a document is assumed to be equivalent to its relatively permanent *aboutness*, but not necessarily to its varying meaning(s), so that one may use the term "*aboutness* analysis" for the cognitive process by which a classifier discerns the inherent at least partially independent topic, subject, or *aboutness* of a document. "*Aboutness* analysis" is preferred to the traditional term "subject analysis" because the latter has been used to describe both *aboutness* and meaning, but the term *aboutness* analysis preserves and emphasizes this fundamental, although unproven, distinction. (p.86)

Plus récemment, Hjørland (2001) émet la proposition que ces concepts devraient être considérés comme synonymes : "Therefore, in my opinion, "subject" and "aboutness" should be considered real synonyms in information science ..." (p.774).

Évidemment, il existe une relation étroite entre l'*aboutness* d'un document et ses significations potentielles pour les lecteurs, la distinction des deux ne devant pas être prise de manière stricte et rigide. Toutefois, cela permet de spécifier qu'un document peut avoir un seul *aboutness*, mais un nombre illimité de sens.

D'ailleurs, Beghtol (1986) rappelle que les systèmes de classification bibliographique sont fondés sur l'entente implicite d'un *aboutness* relativement permanent des documents. La notation et la représentation conceptuelle de ces systèmes ont été élaborées de manière à rendre compte de tous les *aboutness* qu'un document peut présenter, sans qu'ils soient altérés lorsqu'un sens particulier est attribué au contenu du document ou lorsqu'une utilisation particulière en est faite.

Dans ce contexte, le problème de la cohérence interindexeurs ne semble pas témoigner en faveur de la théorie d'un *aboutness* inhérent à un document, si l'on suppose que ce que les indexeurs doivent représenter est effectivement cet *aboutness* et seulement cet *aboutness*. On peut alors poser la question suivante: Est-ce que l'indexation consiste à représenter un document de manière à ce qu'il réponde à un ensemble de requêtes, ou si l'indexation consiste à représenter l'information contenue dans un document?

Comme le souligne Robertson (1979), la fonction d'un système de repérage d'information n'est pas de répondre à des questions, de satisfaire un besoin d'information ou de rectifier un état de connaissance, mais bien d'assister les usagers des systèmes d'information à faire ces choses. En d'autres termes, l'objectif d'un système de repérage par sujet consiste à retrouver des documents dont les *aboutness* signalés orientent l'usager par rapport aux significations potentielles qu'il peut y trouver.

Plusieurs similitudes recourent les différentes acceptions théoriques de la nature du sujet et les directives dans les normes d'indexation concernant l'identification et la représentation du sujet (ou *propos*) d'un document. En effet, les recommandations de la norme ISO (1985) laissent sous-entendre que les indexeurs doivent analyser le *propos* d'un document selon toutes ces significations différentes. Les indexeurs doivent faire une indexation suffisamment précise et exhaustive qui répond au plus grand nombre de questions potentielles des usagers, tout en permettant de discriminer ce document parmi l'ensemble des documents constituant le fonds documentaire. Ils doivent également s'assurer de respecter les contraintes du langage documentaire utilisé ainsi que les politiques du centre de documentation.

Quelle que soit la norme, elles demeurent toutes relativement vagues quant à la façon de concilier les différentes perspectives d'analyse. D'ailleurs, Chaumier (1988) critique un certain manque de rigueur en soulignant que les normes et politiques sont plus riches en conseils de mise en forme qu'en recommandations sur l'analyse de contenu. L'indexeur est laissé plus ou moins à lui-même pour décider de l'information qu'il transposera dans son indexation. Même les organismes possédant des politiques rédactionnelles élaborées ne semblent pouvoir couvrir le flou entourant cette tâche: "C'est la responsabilité des indexeurs de pondérer l'importance des concepts lors de l'indexation" (Bonville, 1983 in Patch-Neveu, 1989, p.8). Parallèlement, Crystal (1995) soulève le paradoxe de la prise en compte de l'utilisateur par l'indexeur:

Indexers are in effect trying to provide answers to a host of unasked questions – and in itself is an interesting reversal of communicative priorities.... In most cases, they do not know who the audience will be; and they do not receive any feedback as to whether their judgements have been successful. From a communicative point of view, there is probably no more isolated intellectual task than indexing. (p. 153)

Quelle que soit l'acception de la nature du sujet d'un document, Beghtol (1986) souligne que très peu de recherches posent clairement la question de la démarche cognitive suivie par l'indexeur pour arriver à identifier correctement cet *aboutness* et/ou ce *meaning*. Albrechtsen (1993) insiste également sur ce point: "the majority of the literature on subject indexing concentrates on step two<sup>14</sup> and fails to provide precise rules for realizing step one where the challenge presented is: finding the subject(s) of a document" (p.219). En somme, on sait peu de choses sur les procédures intellectuelles des indexeurs leur permettant de déterminer ce qui est ou non *indexable*.

Milstead (1994) rappelle entre autres, que l'indexeur doit décider de ce dont le document parle (*aboutness*) et/ou à quelles questions cela répond, pour ensuite traduire cette décision en termes d'indexation. Toutefois, l'auteur remarque qu'il n'y a que très peu d'indices sur les processus mentaux impliqués lorsqu'un indexeur décide du *propos* d'un document et des requêtes auxquelles son indexation permet de répondre (Milstead, 1994).

Pour Milstead (1994), le constat de l'incohérence interindexeurs trouve sa source en partie dans la perception différente qu'ont les indexeurs du *propos* d'un document ou de la valeur estimée de cette information pour les usagers. En somme, elle suppose qu'une fois les variables d'expertise en indexation, d'expertise dans l'utilisation du langage documentaire, de la connaissance du domaine et de la connaissance des politiques du centre de documentation ont été prises en compte, les différences qui persistent résident dans les expériences cognitives différentes des indexeurs avec les documents (Milstead, 1994).

En fait, les chercheurs s'entendent pour dire que l'atteinte d'une cohérence interindexeurs parfaite est un problème pratiquement insoluble qui laisse probablement poindre un problème plus fondamental. Entre autres, cela permet de constater le manque de connaissances concernant les processus intellectuels qui gèrent l'activité des indexeurs:

Yet it is true to say that this most crucial component of the activities studied by information science has been greatly neglected.... An understanding of the processes of indexing and abstracting will not, of course, lead automatically to any improvement in the quality of indexes and abstracts; but it is certainly arguable that if information science is to make any genuine theoretical advances it must seek to understand this most central activity of any information system. (Hutchins, 1977, in Endress-Niggemeyer, 1989, p.230)

Quinze ans plus tard, la situation ne semble guère avoir évolué:

A productive step forward with indexing research would be to begin by examining how indexers actually undertake the process of subject analysis and to explore systematically factors that guide and influence this process. (Todd, 1992, p.101)

De plus, le développement des nouvelles technologies de l'information et des communications (NTIC) au cours des années '90 a fait ressortir l'importance de centrer les efforts de recherche sur le travail intellectuel des indexeurs. De fait, l'expansion rapide de l'Internet et le nombre croissant de bases de données électroniques accentuent la nécessité d'une indexation de qualité, tout en soulevant la complexité grandissante et les difficultés reliées à une telle entreprise (Bates, 2002; Lancaster, 2003; Lunin et Fidel, 1994). Pour plusieurs chercheurs (Milstead, 1994, 2000; Suchman, 1995; Todd 1992), le fait que l'intérêt actuel en SI est au développement des méthodes d'indexation automatisée plutôt

---

<sup>14</sup> *Step one* renvoie à la première étape de l'indexation étant l'extraction des notions essentielles du document, et *step two* renvoie à la deuxième étape, soit la transposition de cette représentation dans les termes d'un

qu'à l'amélioration de l'indexation humaine contribue à la persistance des problèmes de repérage de l'information. Par exemple, l'attribution automatique de mots-clés utilisée pour indexer les documents sur Internet contribue à un repérage relativement inefficace<sup>15</sup>. Même une requête formulée avec des mots mal orthographiés peuvent repérer des notices<sup>16</sup>. Il n'est pas rare d'obtenir des centaines, voire des milliers de références en réponse à une question. Il incombe alors à l'utilisateur de débroussailler cette montagne de références.

Mais comme le souligne Bates (1998), il arrive fréquemment que les références les plus pertinentes pourront échapper à l'utilisateur, ce dernier n'ayant pas utilisé dans sa requête, les mots avec lesquels le document a été indexé. À titre d'exemple, supposons que quatre personnes s'intéressent à la liberté d'expression sur l'Internet et formulent leur requête de manière différente<sup>17</sup> (Bates, 1998):

- "liberté d'expression" + Internet ("*freedom of speech*" + *Internet*)
- "Premier amendement" + Web ("*First amendment*" + *Web*)
- "libre expression" + cyberspace ("*free speech*" + *cyberspace*)
- "liberté intellectuelle" + Net ("*intellectual freedom*" + *Net*)

Les résultats démontrent qu'il n'y a aucun recouvrement parmi les notices retrouvées pour ces requêtes. Précisons que la comparaison est établie à partir des dix premières références repérées pour chacune des requêtes:

Users may simply not realize that the 300 hits they get on a search – far more than they really want anyway – are actually a small minority of the 10,000 records available on their topics, some of which may be far more useful to the user than any of the 300 actually retrieved. (Bates, 1998, p.1190)

De plus, Lawrence et Giles (1998) ont calculé que les meilleurs moteurs de recherche ne couvrent pas plus du tiers de l'ensemble des pages du Web. Avec une telle gestion de l'information, MacDougall (1996) estime que "indexing and searching has gone back to

---

langage documentaire donné.

<sup>15</sup> Rappelons que l'efficacité d'un système d'information dépend en grande partie de la qualité de l'indexation, une mauvaise indexation étant la cause de 90 % des problèmes reliés au repérage. Mais les problèmes de repérage proviennent également d'une imprécision dans l'interrogation du système en rapport avec, par exemple, le degré d'expertise de l'utilisateur dans le domaine de sa recherche et dans l'interrogation, le travail du bibliothécaire consultant, le langage d'interrogation (Bertrand, 1993). Chacun de ces aspects est largement couvert dans la littérature en SI, mais pour les besoins de la présente recherche nous nous limitons à un survol de la distinction indexation humaine versus indexation automatique.

<sup>16</sup> Un exemple de Bates (1998) illustre ce fait, le titre des quatre premières notices repérées comprenait le terme "chidlren" plutôt que "children".

square one, resulting in high retrieval sets and low precision. Searchers seem tolerant of inefficient retrieval provided that access is free" (p.281).

Plusieurs auteurs (Bates, 2002; Lancaster, 1998; Milstead, 1994) rapportent que les systèmes d'indexation automatisée actuels utilisent des méthodes statistiques, syntaxiques et sémantiques qui tentent d'imiter les résultats produits par l'indexation humaine, mais leur succès demeure relativement limité étant donné le manque de connaissances des opérations intellectuelles sous-jacentes au travail des indexeurs. D'ailleurs, l'indexation humaine reste privilégiée dans la plupart des centres de documentation spécialisés puisque l'indexeur humain parvient à une compréhension globale du document et peut tenir compte des besoins des usagers potentiels, ce que ne peuvent faire les systèmes automatiques dont l'analyse se limite principalement aux indices de surfaces (fréquence d'occurrences de chaînes de caractères) (Bertrand-Gastaldy, Giroux, Lanteigne et David, 1994).

Le Tableau IV permet de comparer les différences de représentation du sujet d'une indexation humaine et automatisée. Albrechtsen (1993) propose trois représentations du sujet: simple (*simplistic*), orientée en fonction du contenu (*content-oriented*) et orientée en fonction des requêtes (*requirements-oriented*). Ces types de représentations rendent compte de l'information explicite et implicite du document selon la méthode d'indexation utilisée.

La représentation orientée-contenu nécessite fréquemment une interprétation du document qui va au-delà de sa structure de surface (lexicale ou grammaticale), impliquant la reconnaissance de notions qui ne sont pas explicitement énoncées, mais qu'un indexeur humain saisit rapidement. Cette capacité de reconnaître et d'assigner une notion implicite requiert une analyse d'un niveau d'abstraction plus élevé, ce qui constitue la limite d'une représentation simple. Dans la représentation orientée-requête, un document est considéré comme un *instrument de transmission de connaissance*. L'indexeur ne vise pas en priorité à représenter l'explicite ou l'implicite pour ce type de représentation, mais plutôt à rendre cette *connaissance* accessible, donc visible pour toute personne potentiellement intéressée; cette analyse nécessite également un niveau élevé d'abstraction.

---

<sup>17</sup> Il s'agit ici d'une traduction libre des termes utilisés pour les requêtes citées dans l'article de Bates (1998).

Tableau IV. Interrelation entre les représentations du sujet, les types d'information et les méthodes d'indexation.

Représentation du sujet	Type d'information représentée	Méthode d'indexation
Représentation simple	Information explicite	Extraction
Représentation orientée-contenu	Information implicite	Assigination
Représentation orientée-requête		

*Note.* Tableau extrait et traduit de Albrechtsen, H. (1993). Subject Analysis and Indexing: from Automated Indexing to Domain Analysis. *The Indexer*, 18(4), 219-224 (p.220).

Selon Albrechtsen (1993), l'indexation automatisée procédant par extraction directe des notions énoncées explicitement dans le document est limitée à une représentation *simple*, alors que l'indexation humaine permet de représenter l'information selon chacune des représentations, séparément ou conjointement.

Cette différence dans les possibilités de représentation du contenu des documents soulève le besoin de mettre à profit l'expertise des indexeurs pour améliorer l'efficacité des systèmes d'information (Bates, 2002; Milstead, 1994; Todd, 1992):

There was a period when indexing was often denigrated on the grounds that throwing a search engine at full text was all that was needed. Now the fact that intellectual analysis of resources contributes to effective retrieval is being rediscovered ... (Milstead, 2000, p.14)

Des règles précises ou des principes directeurs de l'activité d'indexation pourraient être extraits de l'expertise des indexeurs et transposés en algorithmes ou systèmes de production afin que les systèmes d'indexation automatisée fassent un traitement de l'information comparable à l'analyse humaine.

La mise à jour des règles internes qui guident les décisions des indexeurs fournira une meilleure compréhension de ce qui constitue la représentation du sujet pour l'activité

spécifique d'indexation, en plus de fournir un cadre nouveau pour l'explication de la problématique de la cohérence interindexeurs. Mais pour mettre à jour l'expertise de l'indexeur, ceci sous-tend que l'on étudie de plus près le travail intellectuel réel des indexeurs, lequel est peu pris en compte dans les écrits en SI:

While library users are obviously crucial actors for the enterprise of developing digital libraries, we are ... recentering the discussion on another field of relevant actors, who often seem lost from view. These are the workers who must do the scanning, structuring, indexing, cataloguing, and document management work implied if visions of the digital library are to become real. (Suchman, 1995, p.15)

Pour combler cette lacune théorique importante, plusieurs chercheurs (Beghtol, 1986; Bertrand, 1993; Bertrand et Cellier, 1995; Bertrand, Cellier, et Giroux, 1994; Bertrand, Cellier, et Giroux, 1996; Endres-Niggemeyer, 1989, 1990, 1994; Farrow, 1991, 1994, 1995, 1996; Tomonori, 1983) ont choisi d'examiner l'indexation sous un nouvel angle et en conséquence, se sont tournés vers les théories et méthodes des sciences cognitives pour tenter d'élaborer un modèle des processus intellectuels impliqués dans l'indexation.

L'indexation étant avant tout une tâche hautement intellectuelle (manipulation de concepts, représentation, abstraction, etc.) prenant place dans un contexte de production régi par plusieurs contraintes, il nous apparaît également adéquat d'examiner les processus cognitifs des indexeurs, d'une part, afin de fournir un modèle explicatif du constat d'incohérence, et d'autre part, afin de mieux comprendre cette tâche fondamentale pour le fonctionnement des systèmes d'information. Il est donc nécessaire de présenter les recherches ayant privilégié cette approche pour situer notre étude en rapport avec elles et montrer en quoi notre perspective s'en inspire et s'en distingue.

## **2. Approche cognitive à l'étude de l'indexation**

Les sciences cognitives se proposent de comprendre les fondements du savoir humain. Pour ce faire, elles font appel à des méthodes empiriques afin de répondre à des questions épistémologiques remontant aux temps des premiers philosophes. Elles visent à expliquer la nature du savoir, ses composantes, ses sources, son développement, etc.

Avant de présenter la perspective théorique choisie, nous devons ouvrir une brève parenthèse en rapport aux nouvelles théories en sciences cognitives, plus spécifiquement la théorie de l'action située, et préciser notre position.

Au cours des années '80 et '90, de nouveaux courants de recherche émergent en sciences cognitives, soit les approches dites de l'action située (Suchman, 1987) et de la cognition distribuée (Hutchins, 1995).

Ces approches s'intéressent principalement à "l'analyse des environnements sociocognitifs dans lesquels les usages prennent place.... Elles placent l'interaction au centre de leur questionnement et visent à saisir le rôle des artefacts techniques dans l'organisation des activités humaines" (Millerand, 2003, p.89).

Plus particulièrement, la théorie de l'action située souligne l'importance d'observer les interactions des humains avec leur environnement en situation réelle, dans leur contexte d'occurrence, postulant l'impossibilité d'étudier toute action hors de son contexte d'émergence. De plus, l'action ne serait pas tributaire d'une forme quelconque de plans, mais plutôt la résultante de réactions et d'ajustements en situation, c'est-à-dire qui surgissent spontanément de l'interaction avec l'environnement (Millerand, 2003).

Conséquemment, l'on comprend bien que cette notion d'action contextualisée s'oppose au paradigme dominant en psychologie cognitive, soit la théorie symbolique de la cognition pour laquelle, les individus agissent en fonction de plans préétablis (Suchman, 1993).

Nous ne pouvons exposer le débat ici, mais précisons toutefois que si pour certains cette opposition préfigure un changement de paradigme en sciences cognitives<sup>18</sup>, pour d'autres, l'action située constitue simplement un nouveau programme de recherche.

---

<sup>18</sup> Un numéro spécial de la revue *Cognitive Science* est consacré à ce débat (*Cognitive Science*, 17, 1993).

D'ailleurs, la position de Simon est très claire à propos de la remise en question par l'action située de la notion de représentation symbolique:

In sum, there is no evidence for cognition without at least minimal representation, and when there is anything problematic about response, the internal representations used to generate the response may be elaborate indeed, and in particular, may incorporate mental imagery. These internal representations constitute an important part of the context, including social context, to which behavior is responsive. Moreover, the internal representations have all the properties of symbol structures. (Vera et Simon, 1993, p.40)

Quel que soit le statut de l'action située au sein des sciences cognitives, l'essentiel ici n'est pas d'exposer ce débat, mais bien de mettre de l'avant les caractéristiques de cette approche pour souligner le fait que nous ne pouvons y recourir pour étudier la problématique de la cohérence interindexeurs selon notre hypothèse de base.

Ainsi, pour comprendre ce qui entraîne des choix différents de descripteurs chez des indexeurs possédant un profil similaire, l'on doit cerner les différences au niveau individuel. Ceci nécessite une analyse préalable de la tâche en fonction des éléments de connaissance utilisés dans l'accomplissement de cette tâche, pour permettre ensuite de saisir les stratégies effectives des indexeurs. C'est pourquoi les objectifs de notre recherche ne sauraient être atteints à travers la théorie de l'action située qui, s'inspirant de l'anthropologie cognitive et l'ethnométhodologie, concentre ses analyses sur les environnements sociocognitifs des activités.

Ainsi, cette approche, tout autant que celle de la cognition distribuée, permettrait de comprendre la tâche d'indexation dans son contexte global de production, c'est-à-dire avec tous les paramètres qui interagissent dans l'environnement quotidien des indexeurs (liens avec les collègues, échanges informels avec les collègues, indices des stratégies de recherche des usagers soulignés par la préposée à la clientèle, satisfaction des usagers réguliers, quota à rencontrer, contraintes de productions, ajustement à la refonte de deux ministères, etc.).

Par ailleurs, les chercheurs de ces courants portent un intérêt marqué aux usages d'artefacts techniques qui, dans notre étude, constituent un aspect relativement secondaire. En effet, si les conséquences du travail des indexeurs pourront affecter les requêtes des usagers ou être prises en compte pour l'élaboration de meilleurs outils de repérage, l'analyse même du travail de l'indexeur ne nécessite pas d'analyser l'interaction avec l'outil technique utilisé, c'est-à-dire la base de données. À ce propos, précisons que les indexeurs n'utilisent pas ou peu la base de données lors de l'indexation, privilégiant papier et crayon lors du choix de descripteurs. Une fois l'analyse complétée, la notice électronique est alors remplie. L'objet central de notre recherche étant justement le processus du choix de termes et descripteurs pour comprendre les différences interindexeurs, l'interaction avec la base de données n'est pas essentiel.

À notre avis, ces récents courants de recherche ne pourraient rendre compte avec précision de l'influence de certaines connaissances sur le choix précis de termes et descripteurs. Il nous faut donc utiliser une théorie qui permette d'avoir accès au niveau de données nécessaire pour atteindre nos objectifs. C'est pourquoi nous nous tournons vers une théorie "classique" en sciences cognitives, soit le modèle du traitement de l'information et plus particulièrement, celle de la résolution de problèmes (Newell et Simon, 1972; Simon, 1973, 1985, 1990).

À l'intérieur des sciences cognitives, le modèle du traitement de l'information cherche à identifier et expliquer les processus mentaux avec lesquels une personne transforme l'information reçue (*input*) pour être en mesure de répondre ou d'agir adéquatement (*output*). Ainsi, les théories sous-jacentes à ce champ de recherches ont été développées afin d'expliquer différents aspects du fonctionnement de la pensée humaine impliqués dans des activités, telles l'organisation de la mémoire, la résolution de problème, la prise de décision, la créativité, etc. :

Le comportement est considéré comme dirigé par un modèle interne de l'environnement construit par le sujet, qui applique ce modèle pour planifier ses actions, prendre des décisions, formuler les messages à transmettre et interpréter les messages reçus. (Johnson-Laird, 1980, in Giroux et Larochelle, 1987, p.10)

Il est important de souligner que les recherches portent essentiellement sur les représentations mentales qui nécessitent un niveau d'analyse propre, faisant abstraction des domaines biologique et neurologique d'une part, et des domaines culturel et sociologique d'autre part (Gardner, 1993).

## 2.1 Compréhension de texte : Modèles de van Dijk et Kintsch

En SI, les sciences cognitives ont été utilisées principalement pour étudier les interactions humain-ordinateur et les comportements de recherche de l'information (*information-seeking behavior*) afin de développer des interfaces de recherche. Par contre, la modélisation des processus d'indexation n'a reçu qu'une attention mitigée et conséquemment n'a suscité que très peu de recherches expérimentales (Farrow, 1994). Les recherches qui ont examiné le traitement analytique (classification, indexation, activité résumante) à l'aide des sciences cognitives se sont inspirées principalement des théories de compréhension de texte de Kintsch et van Dijk (1978) et van Dijk et Kintsch (1983) se situant dans l'axe du traitement humain de l'information.

Plusieurs modèles de compréhension de texte sont proposés dans la littérature, mais pour les objectifs de notre étude, nous nous limitons à une présentation succincte des modèles de Kintsch et van Dijk (1978) et van Dijk et Kintsch (1983) qui sont les modèles auxquels Beghtol (1986), Bertrand (1993), Endres-Niggemeyer (1990, 1994) et Farrow (1991, 1994, 1995, 1996) se réfèrent.

Selon Kintsch et van Dijk (1978), la compréhension d'un texte provient de notre capacité à réduire la totalité d'un texte en des proportions manipulables et conséquemment faciles à mémoriser. Pour construire une telle représentation, deux types de processus sont en continuelle interaction :

- Les processus ascendants (*bottom-up*) fondés sur les données textuelles (*data-driven*) ou le contexte linguistique;
- Les processus descendants (*top-down*) fondés sur les concepts en mémoire (*concept-driven*), c'est-à-dire, le contexte extra-linguistique ou extra-textuel.

Nous verrons plus loin comment ces deux processus sont associés aux macrostratégies textuelles et contextuelles respectivement (van Dijk et Kintsch, 1983).

En d'autres termes, la compréhension d'un texte nécessite simultanément l'intervention des connaissances préalables du lecteur (*top-down*) et des éléments du texte (*bottom-up*), lesquels fournissent les données brutes servant à la construction d'une représentation du texte (macrostructure).

### 2.1.1 *Microstructure et macrostructure*

Le principe fondamental du premier modèle de van Dijk et Kintsch (1978) consiste à prendre la proposition comme unité d'analyse de la signification d'un texte. Dans cette théorie, la structure de surface du texte, sa structure sémantique et les informations stockées en mémoire sont toutes constituées d'unités propositionnelles.

La signification d'un texte peut être décrite comme composée de deux niveaux: une microstructure et une macrostructure. La construction de ces deux structures s'effectue simultanément au niveau macro et micro.

La microstructure est construite à partir du microniveau, c'est-à-dire phrase par phrase. Le lecteur établit des liens entre les idées individuelles du texte, les micropropositions, qu'elles soient explicites ou implicites. La construction de la macrostructure se situe au niveau des séquences de propositions (paragraphe, ensemble de paragraphes), appelées macropropositions, qui représentent les idées globales du texte auxquelles sont reliées toutes les propositions de plus bas niveau. Précisons que la formation de la macrostructure provient d'une généralisation des propositions de la microstructure par la suppression de propositions non pertinentes ou redondantes.

Généralement, ce qu'un lecteur retient du contenu d'un texte est la séquence des épisodes principaux ou, dans le cas d'un texte expositif, les points majeurs de

l'argumentation et les conclusions. Il est plus rare qu'il retienne le détail de la microstructure, c'est-à-dire la séquence particulière des événements dans un épisode ou les stades précis d'une argumentation, et il est fort peu probable qu'il se souvienne de la structure de surface, à moins que cette dernière soit d'un style littéraire frappant. Une fois sa lecture terminée, le lecteur devrait avoir produit une macrostructure du texte, constituée d'un unique réseau de macropropositions organisées hiérarchiquement.

Toutefois, un réseau unique ne peut être construit que dans la mesure où existe un recouvrement parmi l'ensemble des propositions, c'est-à-dire entre chaque nouvelle proposition traitée et celles retenues en mémoire.

Il arrive que des propositions ne soient pas explicitement formulées dans la structure de surface du texte. Elles doivent cependant être inférées par le lecteur pour que la cohérence de la représentation soit maintenue, surtout si le recouvrement d'arguments repose sur ces propositions. van Dijk et Kintsch (1983) soulignent que le lecteur fera alors appel à sa connaissance de la situation évoquée par le texte pour inférer les liens manquants. Cette construction d'un graphe conceptuel se fait au fur et à mesure de la lecture. Ainsi, une proposition occupant une position inférieure dans la hiérarchie lors de son entrée, peut être replacée à un niveau supérieur et entraîner une réorganisation du réseau selon l'évaluation de la situation à un moment donné. Le recours à des connaissances préétablies, lors de l'interprétation d'un texte, laisse entrevoir l'importance de l'utilisation de stratégies. Cette notion de stratégie, constituant l'élément central du modèle de compréhension de texte de 1983 de vanDijk et Kintsch, est présentée plus loin.

Il est admis qu'un "bon" lecteur est plus adroit qu'un "mauvais" lecteur pour exploiter les régularités et les redondances des textes ainsi que pour utiliser de manière efficace le contexte. Les inférences d'un bon lecteur connaissant le domaine seront plus justes, mais surtout procéderont plus d'un traitement descendant, établissant facilement des liens entre les propositions. À l'opposé, une compréhension axée sur un traitement ascendant se limitera aux détails explicites. Les inférences et les anticipations n'ayant qu'une place restreinte dans ce genre de traitement, le lecteur aura de la difficulté à construire une macrostructure cohérente et précise. Précisons cependant que la compréhension de certains

types de textes recourent à des processus descendants plus spécialisés. Dans ce cas, le bon lecteur n'ayant pas la connaissance requise pour établir les liens entre les propositions, aura tout de même de la difficulté à formuler une macrostructure. Chaque lecteur apporte un bagage différent de connaissances lors de la lecture d'un texte et les éléments appelés pour faire des inférences ne sont pas nécessairement identiques d'un lecteur à l'autre. En conséquence, les macrostructures résultantes pourront varier.

### *2.1.2 Macrorègles*

Dans leur premier modèle, Kintsch et van Dijk (1978) définissent un ensemble de règles qui permettent de construire une macrostructure d'un texte. Les macrorègles sont les opérations avec lesquelles le lecteur peut réduire l'ensemble des propositions d'un texte à un ensemble de macropropositions qui rend compte de l'essentiel d'un texte. Cette construction peut résulter en une hiérarchie de plusieurs macrostructures selon la complexité et la longueur du texte. Notons que la construction des macropropositions s'opère sous le contrôle de schémas qui contraignent l'application des macrorègles afin que la macrostructure du texte de base ne soit pas qu'une vague généralisation sans signification.

On entend par schéma, la structuration des connaissances en mémoire permettant une représentation sémantique ou séquentielle de certains types conventionnels de textes ou discours, tels la structure d'une histoire ou des articles scientifiques. Pour certains types de textes, la structure schématique est plus évidente et stable. Par exemple, la structure schématique des textes narratifs et plus particulièrement celle des contes (Propp, 1968) est identifiable par les composantes (ou cases) suivantes: une mise en situation, un but, une complication, une résolution et une finale, souvent associée à une récompense. Les textes scientifiques (expositifs ou argumentatifs) comportent généralement les éléments suivants: la présentation d'un paradigme, la démonstration d'insuffisances, l'exposé du problème, la proposition de nouvelles hypothèses, la vérification de ces hypothèses, la preuve d'une ou plusieurs hypothèses, les conséquences ou la portée de la solution.

Ces schémas fournissent la forme globale du texte autant au niveau de l'organisation des différentes parties du texte que du contenu sémantique de chacune d'elles. On parle alors de *superstructure*. L'élaboration de macrostructures pour ce type de textes consiste à extraire l'essentiel du contenu sémantique correspondant à chaque case prédéfinie de la superstructure appropriée.

Ainsi, un lecteur ayant suffisamment de données de base (textuelles) pour reconnaître un schéma, applique ce schéma pour évaluer et intégrer les propositions subséquentes dans une case appropriée pour la construction de la macrostructure ou pour rejeter l'information non pertinente. La structure du schéma varie en fonction de la connaissance du lecteur. Plus le lecteur est familier avec le type de texte et le sujet traité, plus le schéma activé est élaboré et précis. En conséquence, la macrostructure résultante devrait également être précise et cohérente.

En plus de servir pour reconnaître les différents types de textes, un schéma peut être utilisé pour organiser le contenu d'une histoire ou de tout autre texte qui est présenté de manière inhabituelle.

Plusieurs études (Schnotz, Ballstaedt, et Mandl, 1981 in van Dijk et Kintsch, 1983) ont démontré qu'à la suite d'une première lecture, l'individu n'arrive pas à faire une macrostructure cohérente, produisant plutôt un texte relativement long contenant toute l'information retenue de façon plus ou moins organisée. Ainsi, une deuxième lecture lui permet de construire une macrostructure cohérente et plus courte, un résumé. Soulignons que pour ces études, les participants étaient avisés de produire un résumé après la lecture du texte, et non avant la première lecture. Évidemment, il faut prendre en considération la nature des textes utilisés pour ces expérimentations et la tâche demandée.

Revenons aux règles qui permettent de traiter les micropropositions afin que la construction d'une macrostructure soit possible. Il existe quatre macrorègles: généralisation, élimination, construction et zéro.

La macrorègle **généralisation** est appliquée lorsqu'une seule proposition sert de substitution à un ensemble, pouvant être qualifiée de superordonnée. Il s'agit en fait de réduire la quantité d'information devant être retenue en condensant les données qui appartiennent à une même classe. Cependant, l'information sélectionnée ne doit pas être de niveau trop général, le niveau le plus approprié étant déterminé principalement par le thème du texte. Ainsi, l'expression "êtres vivants" serait trop générale comme substitution à "baleine et marsouin"; "cétacés" ou "mammifères marins" seraient plus justes. La règle cesse d'être applicable lorsque l'information a été réduite à une seule proposition. Les propositions "Réjean répare la télévision, Réjean répare la sècheuse, Réjean répare le réfrigérateur", seront condensées dans la macroproposition "Réjean répare les appareils électroménagers". Cette dernière est alors intégrée à la macrostructure directement. Notons que les généralisations s'appliquent autant aux objets, aux individus qu'aux actions.

La macrorègle **élimination** consiste à supprimer toute proposition qui ne contribue pas au développement du thème. "Marie a vu une balle bleue" sera retenu comme "Marie a vu une balle". Une proposition est non pertinente si elle n'est pas une condition d'interprétation pour des propositions subséquentes. Supprimer renvoie ici à la notion de transfert de la mémoire active ou mémoire à court terme, à la mémoire à long terme. En effet, une information peut être rappelée et réintégrée ultérieurement à la macrostructure si le lecteur s'aperçoit de l'importance inattendue d'un fait.

La macrorègle **construction** implique qu'une séquence de micropropositions soit représentée par une macroproposition correspondant à une situation pour laquelle il existe un scénario (*script*) prédéfini. Le lecteur doit faire appel à sa connaissance du monde pour effectuer cette construction. Par exemple, la séquence "Denis a acheté du bois, du sable, du ciment. Il fait une fondation, érige des murs, un toit et pose des fenêtres" sera réduite à la proposition "Denis construit une maison". Encore une fois, il s'agit de condenser la quantité d'information à retenir en conservant l'essentiel du texte, ce qui est dénoté par les composantes ou les conditions des micropropositions.

Une autre macrorègle peut-être identifiée lorsqu'il y a sélection directe d'une proposition dans la macrostructure, sans réduction ou transformation. Dans ce cas, toutes les informations sont jugées pertinentes. C'est la règle **zéro**.

Le premier modèle de compréhension de texte de Kintsch et van Dijk (1978) met en valeur les opérations sous-jacentes au traitement d'un texte pour la compréhension et la production d'une macrostructure élaborée à partir du rappel du texte. Il a été décrit comment les éléments d'un texte sont organisés en un tout cohérent et résumés à l'essentiel par l'application d'un ensemble de règles. En effet, les macrorègles fournissent un ensemble de procédures simples et cohérentes pour la réduction d'un texte à ses éléments fondamentaux. Toutefois, les macrorègles sont tributaires d'une lecture complète du texte et n'offrent qu'une vue partielle de l'ensemble des dimensions impliquées dans le processus de compréhension.

### 2.1.3 Macrostratégies

Ainsi, pour pallier au caractère rigide des macrorègles, van Dijk et Kintsch (1983) proposent les *macrostratégies*. Ce nouveau modèle de compréhension de texte est fondé sur le constat qu'un lecteur n'a pas besoin d'attendre la fin d'un paragraphe ou d'un chapitre pour pouvoir inférer le propos du texte. Il est alors primordial d'expliquer comment le lecteur arrive à faire des hypothèses judicieuses à propos de ce dont traite un texte et à inférer rapidement une macrostructure à partir d'un minimum d'informations textuelles. La notion de stratégies possède l'avantage d'être moins dépendante d'un traitement ascendant et met l'accent sur les processus plutôt que sur des règles, ce qui lui confère une capacité d'adaptabilité à une variété de situations. Les deux perspectives ne s'excluent pas mais démontrent bien l'apport complémentaire et nécessaire des traitements ascendants et descendants, privilégiant les premiers pour l'application de règles et les derniers pour l'utilisation de stratégies.

van Dijk et Kintsch (1983) regroupent les macrostratégies en deux classes: les macrostratégies contextuelles et les macrostratégies textuelles.

### Macrostratégies contextuelles

L'utilisation des stratégies contextuelles accélère et facilite la construction de la macrostructure puisqu'elles permettent au lecteur de faire de fortes inférences sur le thème du texte et les différents types d'informations pouvant y être exprimés.

Un type de macrostratégies fait appel aux connaissances générales ou connaissance du monde (*world knowledge*) du lecteur. Bien que tous les textes ne sont pas prévisibles au même niveau, les thèmes ou le déroulement de certains textes peuvent être plus ou moins stéréotypés. Ainsi, le lecteur exploite les régularités dans les événements et dans les actions rencontrés dans le quotidien comme éléments de référence pour anticiper la macrostructure d'un texte traitant de ces mêmes événements ou actions. Cette anticipation repose sur le postulat que les régularités macrostructurelles des faits réels et des discours de ces mêmes faits seront similaires.

Un deuxième type de macrostratégies utilisé par le lecteur pour inférer la macrostructure d'un texte se rapporte aux connaissances relatives à certains types de textes. Le lecteur connaît le type d'actions ou d'événements décrits par un genre de textes. Pour les "nouvelles" de l'actualité, le lecteur s'attend à lire à propos d'événements politiques ou d'événements importants telles une catastrophe naturelle, la découverte d'une planète, etc., mais ne s'attend pas à la description d'événements quotidiens triviaux. Ceci sous-tend que la plupart des types de textes et de discours sont contraints à un ensemble de thèmes possibles (*topic set*). De toute évidence, cet ensemble est relatif et n'est pas déterminé uniquement par le type de textes, mais également par des normes socioculturelles, des valeurs, des intérêts, etc.

van Dijk et Kintsch (1983) ont dressé une liste des informations contextuelles utilisées dans les macrostratégies pour inférer les thèmes ou les ensembles de thèmes (*topic*

sets) d'un texte ou d'un fragment de texte (Tableau V). La liste comporte trois catégories qui s'emboîtent l'une dans l'autre, la première étant la plus générique<sup>19</sup>.

Tableau V. Informations contextuelles utilisées pour les macrostratégies

- 
1. *Culture générale (world knowledge)*
    - a. Activités générales, objectifs d'un groupe culturel (ex. agricole)
    - b. Événements ou actions spécifiques (ex. rites)
    - c. Circonstances biophysiques spécifiques (ex. climat, paysage, animaux, plantes)
    - d. Objets spécifiques (ex. produits industriels)
  2. *Situation socioculturelle (partie de 1)*
    - a. Type de situation (ex. déjeuner, trajet en autobus, mariage)
    - b. Catégories de participant
      - (i) Fonctions (ex. conducteur d'autobus, docteur, juge)
      - (ii) Rôles (ex. mère, ami)
      - (iii) Caractéristiques sociales (ex. sexe, âge, etc.)
      - (iv) Caractéristiques individuelles (ex. personnalité, intérêts, objectifs)
    - c. Événements ou interactions typiques (ex. aider, payer, consulter)
    - d. Conventions (ex. lois, règles, habitudes)
  3. *Situation de communication (partie de 2)*
    - a. Objectifs globaux de l'interaction
    - b. Actes de langage globaux et locaux
    - c. Contexte référentiel (ex. présence des personnes, objets)
- 

*Note.* Traduction libre du tableau présenté dans van Dijk, T.A., Kintsch, W. (1983). *Strategies of Discourse Comprehension*. London: Academic Press (p.199).

La première catégorie concerne les connaissances associées à la culture générale. Elle comprend les activités générales; les objectifs d'un groupe culturel donné, par exemple, l'agriculture; les événements spécifiques, tels les rites. La deuxième catégorie spécifie les dimensions socioculturelles, tels les types de situations comme le déjeuner; les catégories de participants caractérisés par leur fonction, leur rôle, leurs caractéristiques sociales et individuelles; les interactions et les événements typiques comme aider; de même que les conventions comme les lois, les règles, etc., qui sont à leur tour spécifiées par les dimensions de la troisième catégorie. Ce sont les situations de communication qui regroupent les objectifs de l'interaction, les actes de langage et le contexte référentiel.

Pour comprendre un texte, le lecteur doit posséder un certain ensemble de connaissances se rapportant à plusieurs catégories qui sont communes à celles de l'auteur.

---

<sup>19</sup> Rappelons que la théorie de van Dijk et Kintsch (1983) couvre tous les types de discours, autant oraux qu'écrits, et que nous avons cependant limité notre couverture aux textes écrits qui constituent l'un des objets de notre étude.

Si ce n'est pas le cas, une compréhension partielle du texte en résultera. En d'autres termes, aucune macroproposition ne pourra être formée sans le recours à la connaissance de concepts de niveau élevé.

### Macrostratégies textuelles

Si les macrostratégies contextuelles créent des attentes chez le lecteur à propos du thème d'un texte, la décision finale pour l'identification du véritable thème du texte dépend des caractéristiques même du texte. Les stratégies contextuelles et textuelles vont nécessairement de pair.

Van Dijk et Kintsch (1983) identifient cinq types de stratégies textuelles: les stratégies utilisant les indices structuraux et celles utilisant les marqueurs de changement de propos, les stratégies syntaxiques, les stratégies sémantiques et les stratégies schématiques.

### **Indices structuraux**

En principe, l'utilisation d'une macrostratégie contextuelle suppose que les macropropositions n'ont pas besoin d'être présentes à la surface du texte, mais peuvent être inférées à partir d'une interprétation de la sémantique des mots, des phrases et des séquences de phrases. Cependant, pour faciliter de telles inférences, les textes annoncent normalement la macrostructure projetée par des indices structuraux. Afin de limiter les interprétations liées aux connaissances personnelles, le texte mentionne souvent de manière explicite les macropropositions, c'est-à-dire les thèmes principaux. Ces expressions thématiques peuvent apparaître à divers endroits, mais se situent normalement à des endroits stratégiques, soit en début ou fin de texte, en début ou fin de paragraphe ou de tout segment pertinent.

La position occupée par ces expressions témoigne de leur fonction sémantique. Placées en début de texte ou de segment, elles aident le lecteur à inférer une macroproposition et à interpréter les propositions subséquentes en fonction de cette

hypothèse. En fin de texte ou de segment, les expressions thématiques servent à vérifier, rappeler ou corriger la macroproposition existante. Elles pourront également être signalées par une typographie différente et être séparées des autres expressions par des interlignes. L'auteur pourra utiliser une ponctuation particulière pour souligner les thèmes centraux. De même, la longueur de ces expressions varie selon qu'elles constituent un résumé (une séquence de phrases), un titre ou un sous-titre (une proposition, un mot). En général, seuls les résumés expriment les macrostructures, tandis que les titres et sous-titres expriment uniquement les macropropositions de haut niveau.

Ainsi, les expressions thématiques ont une double fonction. D'une part, elles soulignent les concepts importants pour aider le lecteur à former la macrostructure et d'autre part, elles activent les connaissances pertinentes du lecteur pour l'interprétation des propositions subséquentes et l'application des macrorègles. D'ailleurs, il a été démontré qu'un lecteur peut identifier les informations importantes d'un texte sans nécessairement comprendre les idées principales ou l'argumentation de l'auteur (Rossi et Bert-Erboul, 1991 in Bertrand, 1993).

### **Stratégies syntaxiques**

Les indices syntaxiques se situent au niveau de la microstructure et indiquent l'importance locale d'un terme ou d'une expression. Ils permettent d'établir les liens entre les différentes propositions, assurant le maintien de la cohérence. Ces stratégies n'agissent qu'indirectement sur la construction de la macrostructure par le biais de l'accumulation d'inférences issues du microniveau. Par exemple, si le mot "John" revient en tant que sujet de plusieurs phrases et occupe une position thématique, le lecteur assumera qu'il est le thème de séquences, ce qui entraînera probablement l'inférence qu'il est l'acteur principal du texte, donc un agent de la macroproposition.

### **Marqueurs de changement de propos**

Les marqueurs de changement de propos soulignent le passage d'un épisode à un autre. Un épisode correspond à une séquence de phrases résumées par une seule

macroproposition. Puisque chaque épisode est résumé ou représenté par une macroproposition différente, un élément différent sera introduit au début d'un nouvel épisode. van Dijk et Kintsch (1983) donnent quelques exemples de marqueurs de changement de propos en début d'épisode: marqueurs de changement de temps, de période, de lieu, de point de vue, ainsi que l'introduction de nouveaux acteurs. L'un des marqueurs de changement de propos les plus fréquents est le retrait de paragraphe.

Une autre catégorie de marqueurs est identifiée, soit les macroconjonctions retrouvées dans les expressions qui introduisent une nouvelle macroproposition et assurent le suivi de la cohérence pour la formation de la macrostructure. On les retrouve dans les expressions, tels, mais, cependant, au contraire, de plus, etc.

L'utilisation de cette stratégie permet au lecteur de reconnaître qu'une phrase ne peut pas être représentée dans la dernière macroproposition et de créer une nouvelle macroproposition qui intègre les marqueurs de changement de propos repérés.

### **Stratégies sémantiques**

Les stratégies soulignées par les marqueurs de changement de propos ont d'ores et déjà exposé certaines stratégies sémantiques. En effet, bien que les macropropositions puissent être exprimées directement par les indices de surface, elles ne peuvent être inférées correctement qu'à partir d'une représentation de leur signification. C'est donc la signification des mots, des énoncés et des phrases qui fournissent les indices essentiels pour l'inférence des thèmes principaux d'un texte. La formation de macropropositions et la construction de la macrostructure d'un texte sont basées principalement sur l'utilisation de stratégies sémantiques. Ainsi, un lecteur infère la macroproposition pertinente d'un passage dès que l'interprétation de quelques éléments (une phrase, une proposition, un mot) le lui permet. Sur la base de ces macropropositions inférées, il élabore des hypothèses sur les thèmes développés qui lui serviront pour l'interprétation des propositions subséquentes, ces dernières servant également à confirmer ou infirmer les hypothèses.

L'efficacité de ces stratégies sera d'autant plus grande que les connaissances générales du lecteur, à propos des situations mentionnées dans le texte, sont élaborées. Effectivement, ces connaissances sont fondamentales pour l'inférence de macropropositions. Ainsi, si un événement, une action ou un état décrit dans le texte correspond à un scénario (*script*) connu du lecteur, il lui sera plus facile d'inférer une macroproposition en appliquant la structure générale de ce même scénario ou une partie de ce scénario sur le texte.

### **Stratégies schématiques**

Finalement, en plus de la structure de surface et de l'information sémantique, les informations associées à la superstructure d'un texte contribuent à la formation de macropropositions de manière stratégique. Ceci vient du fait que plusieurs structures schématiques, telles les structures narratives, ont un ordonnancement canonique et les catégories schématiques sont gérées par des contraintes sémantiques globales. Par exemple, il est admis que le début d'un récit est schématiquement constitué d'une mise en situation et donc que les premières macropropositions pourront dénoter une description d'état, l'introduction des participants, la spécification du temps et des lieux, ainsi que les circonstances ou les motivations en fonction des événements ultérieurs. Ce que nous avons mentionné précédemment à propos des schémas tient également pour cette macrostratégie. Ainsi le lecteur peut anticiper la structure et le contenu de certains types de textes à partir de ses connaissances acquises de ces types de textes. Une expérimentation de Dillon (1991) a permis de constater que "les lecteurs expérimentés dans l'utilisation de certains types de textes possèdent une superstructure de ce texte leur permettant de prédire avec un haut degré d'exactitude où l'information est localisée" (Dillon, 1991 in Bertrand, 1993, p.71).

Les stratégies schématiques permettent non seulement une représentation structurale d'un texte, mais une représentation fonctionnelle. Tout comme les macrostratégies précédentes, les stratégies schématiques entretiennent un lien d'interdépendance avec les macrostratégies sémantiques.

La description des macrostratégies rend compte du rôle prédominant de la notion d'inférence dans ce deuxième modèle de compréhension de texte de van Dijk et Kintsch (1983). En effet, nous avons pu constater que ce sont les stratégies sémantiques qui unissent les informations de surface aux informations superstructurelles ou contextuelles et permettent au lecteur de construire une macrostructure cohérente du texte. Il a été mentionné également que la diversité des informations utilisées pour l'élaboration de la macrostructure conduira probablement différents lecteurs à produire des macrostructures différentes pour un même texte, selon leurs connaissances générales (*world knowledge*) et leurs connaissances plus spécifiques liées à leur contexte socioculturel, leur fonction, leur intérêt, etc.

Cette dernière remarque n'est pas sans rappeler les multiples facteurs identifiés précédemment comme sources de variations interindexeurs pour la production de listes de termes d'indexation. Mais d'autres parallèles, entre les tâches de compréhension de texte et d'indexation, ont été observés et exploités par plusieurs auteurs en SI afin de fournir des bases théoriques pour le développement d'un modèle de la tâche d'indexation qui rende compte des processus mentaux impliqués pour cette tâche. Ces recherches permettent également d'expliquer de manière indirecte certains aspects du manque de cohérence interindexeurs.

## 2.2 Recherches sur l'indexation en tant que compréhension de texte

### 2.2.1 *Beghtol*

Beghtol (1986) est la première auteure en SI à relever la pertinence des modèles de compréhension de texte de van Dijk et Kintsch pour l'explication des processus intellectuels impliqués lors de la production de textes secondaires. Les bases théoriques proposées par Beghtol assument que ce sont les mêmes processus qui sont nécessaires à l'accomplissement de la tâche d'indexation que ceux définis pour la compréhension de texte. En conséquence, sa théorie demeure relativement générale et colle de très près au modèle original de van Dijk et Kintsch. Cependant, son étude fait ressortir certaines

particularités de la tâche d'indexation qui devront être examinées attentivement pour l'élaboration d'un modèle plus précis.

Entre autres, Beghtol (1986) souligne qu'une théorie des opérations intellectuelles pour l'analyse documentaire devrait expliquer comment l'indexeur s'y prend pour opérer la séquence suivante :

- Prendre la structure de surface du document et la transposer en une structure profonde pour être en mesure d'avoir une représentation la plus juste possible des concepts abordés;
- Prendre la structure de surface des termes autorisés (le vocabulaire du langage documentaire) et la transposer également en une structure profonde;
- Faire concorder les deux structures profondes, c'est-à-dire la structure profonde de ce que l'indexeur croit être la représentation sémantique du document avec la structure profonde des termes potentiels du langage documentaire. Tout ceci dans le but de vraiment faire ressortir la "sémantique essentielle" du document de la façon la plus appropriée, à travers la structure de surface du langage documentaire.

Cependant, Beghtol (1986) nous met en garde contre un biais potentiel introduit lors de la compréhension d'un document en vue de produire un indexat. En effet, il a été démontré que l'objectif de la lecture influence la compréhension du texte, c'est-à-dire que les inférences sont orientées en fonction du but de la lecture (Black, 1981 in Beghtol, 1986). Ainsi, l'indexeur sachant qu'un document doit être indexé à l'aide de termes particuliers d'un langage documentaire, peut accorder une plus grande valeur à certains concepts du document qu'il sait être représentés par ce langage et dévaluer d'autres concepts qui sont plus difficilement traduits par ce même langage.

En d'autres termes, un indexeur sachant que la macroproposition de plus haut niveau d'un document doit être insérée dans une des macropropositions disponibles d'un langage documentaire, ne parviendra probablement pas à faire une analyse objective du document primaire, c'est-à-dire indépendante de sa connaissance du langage documentaire. Cette

éventualité indique que des documents traitant de sujets similaires pourront être indexés différemment par différents indexeurs ou par le même indexeur en des temps différents.

Ainsi, Beghtol identifie théoriquement une des causes du manque de cohérence interindexeurs en relation avec les théories de compréhension de texte et décrit à quel niveau de l'analyse cette incohérence se situe. Selon Beghtol (1986), des études empiriques permettraient de démontrer exactement la manière dont les évaluations différentes de la pertinence des termes d'indexation prennent forme lors de l'analyse du propos de documents primaires. Il est donc justifiable de chercher les causes de l'incohérence dans l'analyse du propos des textes primaires puisque pour Beghtol (1986), "knowledge of causes of inconsistency might lead to methods of lessening inconsistencies or anomalies in document analysis" (p.104).

### 2.2.2 Farrow

À l'instar de Beghtol (1986), Farrow (1991, 1994, 1995, 1996) s'inspire des modèles de compréhension de texte de van Dijk et Kintsch pour fournir un cadre théorique à l'indexation. Non seulement souligne-t-il les parallèles et les différences entre les deux tâches, mais il dresse également un modèle des procédures de compréhension de texte adapté aux exigences spécifiques de la compréhension de texte pour l'indexation.

Il est à souligner que Farrow suggère des principes théoriques s'appliquant aux différents types de traitement analytique: classification, indexation et condensation. En fonction des objectifs de notre recherche, nous limitons la présentation aux éléments théoriques avancés pour un modèle de l'activité d'indexation.

À partir de la description de l'indexation établie par la norme ISO (1985), Farrow (1991) suppose que tout modèle cognitif de l'indexation doit nécessairement rendre compte de deux types de procédures, soit les procédures reliées à la compréhension de texte de l'indexeur, et les celles reliées à la production d'une représentation de ce texte en termes d'indexation.

Ainsi, Farrow (1991) suppose qu'un indexeur comprend un texte essentiellement de la même façon qu'un lecteur, à l'exception de quatre modifications en rapport aux exigences de la tâche:

- 1) Les contraintes de temps reliées à l'indexation ne permettent aux indexeurs qu'un parcours rapide (*scanning*) du document plutôt qu'une lecture à un rythme normal.
- 2) La compréhension de texte des indexeurs est axée principalement vers le but de l'indexation. Ces conditions de compréhension diffèrent grandement de celles rencontrées pour la mesure de la compréhension de texte dans les expériences psychologiques qui ont permis d'élaborer les modèles explicatifs de cette activité. En effet, ces expériences mesurent la proportion d'un texte qu'un individu comprend ou retient après un certain temps.
- 3) La compréhension de texte des indexeurs est immédiatement suivie de la production d'une liste de descripteurs. Un modèle de la compréhension de texte des indexeurs doit alors être relié à un modèle de production approprié.
- 4) La plupart des indexeurs oeuvrent à l'intérieur d'un éventail restreint de types de textes et de domaines. En conséquence, plusieurs éléments sont rencontrés de manière répétitive entraînant un automatisme dans le traitement des textes qui ne se retrouve pas dans des conditions de lecture normale.

Ces premières considérations permettent de souligner que l'application des macrorègles au processus d'indexation est quelque peu problématique, en ce sens que les conditions de production de l'indexation, obligeant un survol (*scanning*) du document, ne correspondent pas à la lecture intégrale que requièrent les macrorègles. Farrow (1996) indique que leur rôle se limitera à la description des opérations de compréhension lorsque l'indexeur s'attarde de manière ponctuelle à un passage particulier, qu'il lit en entier.

D'ailleurs, une autre étude de Farrow (1996) souligne les limites des macrorègles pour expliquer les processus intellectuels impliqués dans la construction d'une macrostructure succincte. Les résultats de son expérimentation démontrent que l'application des macrorègles, à un court passage d'un texte, entraîne la production de courts énoncés

différents en fonction du contexte dans lequel le document est analysé, ainsi qu'en fonction de la connaissance générale du lecteur et de son jugement. De plus, Farrow souligne: "quite radically different reductions would be applied if text were being indexed as part of a work supporting a particular thesis.... With such variations in the analyses of even short, simple passages of texts, what hope can there ever be for inter-indexer consistency?" (p.14).

Farrow tente donc d'être plus spécifique dans son adaptation du modèle de compréhension de texte pour expliquer les processus intellectuels impliqués pour l'indexation en se tournant vers les macrostratégies textuelles et contextuelles. Il considère alors l'indexation comme une activité de lecture rapide avec but, c'est-à-dire orientée en fonction de la tâche à accomplir.

Par exemple, le peu de temps alloué pour le traitement d'un document entraîne nécessairement le développement de stratégies permettant de rendre compte de l'essentiel du document en fonction des conditions de production. Ainsi, les indexeurs suivent une liste plus ou moins stricte des portions de texte qui doivent être considérées avec plus d'attention. Rappelons à ce sujet les recommandations de la norme ISO (1985) concernant l'attention particulière que l'indexeur doit accorder aux éléments suivants : titre, résumé, sommaire, introduction, débuts de chapitres et de paragraphes, conclusion, illustrations, diagrammes, tableaux et leur légendes, mots ou groupes de mots soulignés ou mis en valeur par une typographie différente.

Ainsi, l'indexeur effectue une lecture sélective d'un document en se servant abondamment des macrostratégies textuelles pour le guider dans le repérage d'indices perceptuels utiles pour la reconnaissance des notions essentielles d'un document. D'ailleurs, les écrits en SI ont reconnu depuis longtemps la prise en compte de la structure des textes et des indices perceptuels pour l'analyse d'un document (Jones, 1983; Hutchins, 1975).

Farrow reconnaît donc la pertinence des macrostratégies textuelles pour l'indexation et précise toutefois l'utilisation marquée de certains indices en rapport aux caractéristiques d'une lecture faite dans le but d'indexer. Ainsi, la fréquence d'occurrence d'un mot ou d'une

notion est un facteur important dans la détermination de son *indexabilité*. D'ailleurs, Jones (1983) a identifié que les mots rares et inusités ainsi que les mots d'une longueur inhabituelle augmentent la possibilité d'une sélection. De même, les mots qui apparaissent dans le premier paragraphe d'un texte, les premières phrases de paragraphes et les mots qui sont définis sont plus susceptibles d'être retenus comme termes d'indexation. À l'inverse, il semble que les paragraphes de conclusions ne soient pas un endroit privilégié pour le choix de termes d'indexation.

Farrow émet alors l'hypothèse qu'un indexeur ayant identifié un mot au tout début de son analyse, comme terme potentiel d'indexation, gardera ce mot à l'avant plan de sa mémoire pour la suite de l'analyse et toute autre occurrence de ce même mot contribuera à renforcer cette position. Jones (1976) soutient également ce point de vue en suggérant que l'interaction des processus ascendants et descendants est telle, pour l'indexation, qu'un indexeur ayant remarqué ou retenu un mot aura une forte tendance à s'attarder aux mots ayant une proximité sémantique avec ce dernier. Cette hypothèse peut vraisemblablement s'appliquer au processus d'indexation puisque l'activation (*priming*) des réseaux sémantiques a été démontrée par plusieurs études en psychologie (Collins et Quillian, 1969; Smith, Shoben et Rips, 1974; Collins et Loftus, 1975; Posner et Snyder, 1975 in Fortin et Rousseau, 1989).

Farrow mentionne également que les conditions de production de l'indexation obligent les indexeurs à recourir abondamment aux stratégies sémantiques et schématiques. Ces stratégies seront d'autant plus efficaces chez un indexeur qui travaille avec un nombre restreint de types de documents, ce qui lui permet une reconnaissance rapide de la superstructure. Parallèlement, l'activation des contenus sémantiques se rapportant à ce type de document accélère l'évaluation des termes d'indexation potentiels à partir de la prise de connaissance d'un minimum d'informations du document. De plus, la norme ISO (1985) suggère un ensemble de questions que peuvent se poser les indexeurs pour circonscrire l'essentiel du contenu d'un document. Ceci constitue en soi un cadre conceptuel préétabli pour la compréhension de texte particulière à l'indexation (Farrow, 1994):

- a) Le document traite-t-il d'un objet affecté par une activité?

- b) Le sujet contient-il une notion active (par exemple une action, une opération, un processus, etc.)
- c) L'objet est-il affecté par l'activité identifiée?
- d) Le document traite-t-il de l'agent de cette action?
- e) A-t-il recours à des moyens particuliers pour accomplir l'action (par exemple instruments spéciaux, techniques ou méthodes spéciales)?
- f) Ces facteurs étaient-ils placés dans le contexte d'un lieu ou d'un environnement particulier?
- g) Des variables dépendantes ou indépendantes sont-elles identifiées?
- h) Le sujet a-t-il été traité d'un point de vue particulier que l'on n'associe pas normalement à ce domaine d'étude (par exemple une étude sociologique de la religion)?

Il a été mentionné précédemment que l'indexation sera accomplie de manière plus efficace si l'indexeur possède une certaine connaissance du domaine des documents qu'il doit indexer. De plus, le traitement sélectif de l'information essentielle du document est hautement conceptuel (descendant) lorsque la lecture doit répondre à un but précis. Dans le cas présent, la production d'un indexat est la condition finale de la lecture d'un indexeur et ce qui influence sa compréhension des documents.

Ainsi, Farrow (1991) suggère un ensemble d'opérations intellectuelles effectuées par l'indexeur lorsqu'il évalue la pertinence d'un terme pour la liste finale de termes d'indexation: renforcement, modification, regroupement, rejet.

Mentionnons cependant que ces opérations reposent sur le postulat que l'indexeur cherche à se créer une représentation de l'essentiel du contenu, dès le premier contact avec le document (normalement le titre), et qu'il mettra à profit toutes ses connaissances, autant générales que spécifiques, afin de parvenir à son but. Ceci place alors le traitement descendant au premier plan des processus impliqués dans l'indexation:

- 1) Opération de renforcement (*reinforcement*) : un terme est jugé cohérent avec le propos global du document identifié précédemment et/ou peut être apparié à un terme déjà retenu.
- 2) Opération de modification : un terme est modifié en fonction des connaissances ou de l'objectif de l'indexeur, ou en fonction d'un terme déjà retenu pour la représentation du document.

- 3) Opération de regroupement (*chunking*): un ensemble de termes reliés par leur sémantique sont fusionnés en un seul terme.
- 4) Opération de rejet : un terme est jugé incohérent avec le propos global du document identifié par l'indexeur, ou jugé incohérent en fonction des connaissances spécifiques de l'indexeur ou de son objectif.

Ces opérations demeurent à un niveau de description relativement général. Elles permettent néanmoins de saisir le processus sous-jacent à une continuelle restructuration de la représentation du document par un va-et-vient entre les éléments textuels et la connaissance des éléments contextuels.

Pour pouvoir rendre compte de l'ensemble des éléments impliqués dans la compréhension de texte de l'indexeur, Farrow (1991, 1996) s'inspire également des macrostratégies contextuelles de van Dijk et Kintsch (1983). Rappelons que les trois catégories générales d'information utilisées par les macrostratégies contextuelles proviennent des connaissances issues de la culture générale (*world knowledge*), de la situation socioculturelle et de la situation de communication.

Selon Farrow (1996) les groupes de stratégies contextuelles utilisées pour l'indexation sont les suivants:

- 1) Professionnel : L'application des standards professionnels, notamment de la norme ISO 5963, laquelle comprend des prescriptions générales pour l'analyse de contenu ainsi que pour la prise en compte des critères de qualité à respecter.
- 2) Situationnel : La prise en compte des besoins des usagers du système documentaire pour le choix des termes d'indexation. L'indexation devra demeurer à un niveau relativement général si le système dessert une population de jeunes adolescents et être très spécifique si la clientèle est composée de scientifiques.
- 3) Externes : Les contraintes de temps de production, les contraintes imposées par le langage documentaire utilisé, les exigences des politiques du centre de documentation, tels, le nombre de termes à assigner, le type d'information privilégiée, etc.

Ces macrostratégies sont soutenues par la connaissance générale de l'indexeur en relation avec ses compétences de lecture et d'écriture, sa connaissance du contexte général de la situation présentée dans le document, et de ses connaissances particulières reliées au sujet spécifique du document (Farrow, 1991, 1996).

À ce modèle théorique de compréhension de texte pour l'indexation, Farrow (1991, 1996) superpose quelques principes généraux pour un modèle de production. Tout comme van Dijk et Kintsch (1983), Farrow (1991) reconnaît que les modèles de productions de texte sont plus rares. Ce constat s'explique par la nature abstraite des données initiales pour l'élaboration de modèles de production. En effet, si le texte contient les données de base pour tester la compréhension de texte chez un individu, la production repose sur "une idée vague, une certaine intention, etc." qui sont beaucoup plus difficiles à contrôler et manipuler en conditions d'expérimentation (Farrow, 1991).

En ce sens, Farrow (1991, 1996) suppose qu'un modèle de production, pour l'indexation, devrait être plus simple à élaborer puisque la tâche est hautement structurée et que la production de l'indexat découle directement de la compréhension du document, donc d'un modèle de compréhension.

Le processus de production pour l'indexation génère un ou (normalement) plusieurs termes indiquant les thèmes contenus dans le document. Cette production est gérée par un ensemble de contraintes propres au système documentaire utilisé. Ainsi, l'ampleur des contraintes pour l'indexation varie selon que le langage utilisé par le système est un vocabulaire contrôlé ou libre. Un vocabulaire contrôlé imposera plus de contraintes qu'un vocabulaire libre qui n'est contraint par aucun ensemble de termes, structure ou relation prédéfinis.

Farrow (1991, 1996) présente les modèles de compréhension et de production séparément, mais souligne qu'en pratique, l'un et l'autre se fusionnent, en ce sens qu'un terme d'indexation est à la fois un élément utilisé pour la compréhension de texte et pour la production de la liste de termes d'indexation. Évidemment, les exigences du contexte de la

tâche influenceront également la compréhension; plus le vocabulaire du langage est exhaustif plus la compréhension doit être précise.

Une fois le propos du document identifié, l'indexeur doit comparer sa représentation aux termes du thésaurus et apporter des modifications à sa liste lorsque les termes retenus ne correspondent pas à des descripteurs contenus dans le thésaurus (avec l'utilisation d'un vocabulaire contrôlé).

Farrow (1991, 1996) rappelle que son modèle ne constitue qu'une possibilité pour l'élaboration d'une théorie des processus d'indexation. Il souligne entre autres qu'il est élaboré à partir des résultats de recherches orientées principalement en vue de la compréhension de texte plutôt qu'en fonction d'une production dirigée par un but très précis.

Malgré que l'auteur admet que son modèle est incomplet et qu'il demeure à un niveau de description relativement superficiel, quelques principes émergent :

- L'importance du traitement descendant (conceptuel) et en conséquence, la nécessité pour les indexeurs d'être familiers avec le domaine des sujets abordés par les documents;
- Les limites de la compréhension d'un texte engendrée par le survol (*scanning*) d'un texte et l'utilisation des différents types d'indices perceptuels (ou macrostratégies textuelles) par les indexeurs lors de ce survol;
- La nature et l'importance de l'expertise des indexeurs dans l'accomplissement d'une tâche professionnelle.

Farrow (1991, 1996) conclue en soulignant que, malgré leurs imperfections, les macrorègles et les macrostratégies fournissent l'explication la plus complète qui soit jusqu'à présent de la procédure de réduction d'un texte à ses éléments essentiels. Le besoin de recherches sur les processus d'indexation demeure, ne serait-ce que pour valider de manière expérimentale les parallèles établis entre les modèles de compréhension de texte et la tâche.

Il a été démontré que comparativement aux expériences ayant servi à l'élaboration des modèles de compréhension de texte, l'indexation ne consiste pas à produire un résumé à partir du rappel d'un texte puisque l'indexeur peut consulter le document tout au long de l'indexation. En fait, ce qu'on demande aux indexeurs est de produire une sorte de résumé télégraphique généralement dépourvu de toute syntaxe, composé essentiellement de noms et d'adjectifs représentant l'essentiel du contenu d'un document. Cependant, le fait que l'indexeur doit traduire sa représentation initiale d'un document en termes d'un langage documentaire suppose qu'il doit également identifier et faire ressortir en quoi ce document est différent des autres documents de la base de données qui traitent de propos similaires. Cette distinction vise également à maximiser l'efficacité du repérage. Cela demande donc aux indexeurs d'avoir une expertise non seulement dans le domaine du document indexé, mais aussi une connaissance approfondie des termes autorisés pour traduire la représentation initiale le plus adéquatement possible en fonction de l'ensemble des documents du système documentaire.

Le modèle de Farrow fait ressortir le rôle prédominant qu'occupent les dimensions contextuelles dans la réalisation de l'indexation, si bien que l'on s'éloigne de toute compréhension de texte "pure". Une étude de Tomonori (1983) soulignait déjà l'effet prégnant des connaissances acquises (*world knowledge*) de l'indexeur sur l'ensemble du processus. En effet, comme Beghtol (1986) a souligné que l'indexeur peut biaiser la représentation du propos du document en fonction de sa connaissance du langage documentaire, on peut aisément concevoir que la prise en compte des besoins des usagers, des critères de qualité, des politiques institutionnelles, etc., vont imposer des contraintes aux indexeurs et baliser leurs décisions pour le choix de termes d'indexation.

Cependant, aucun des modèles ou principes théoriques proposés par ces auteurs ne dit dans quelle mesure chacune des dimensions contextuelles influence les différentes étapes de l'indexation et surtout, de quelle manière les indexeurs en tiennent compte pour choisir leurs termes d'indexation. En conséquence, ils peuvent encore moins expliquer les différences individuelles de choix de termes. Évidemment, ces modèles visent à expliquer de manière générale l'ensemble des opérations intellectuelles en jeu lors de l'indexation, et non pas à expliquer les différences individuelles à l'intérieur de ce modèle global.

D'ailleurs, les modèles dont ils s'inspirent pour élaborer leurs théories ne permettent pas de rendre compte des différences individuelles : "this abstract, structural description of macrounderstanding could hardly provide a sound explication of individual differences and differences in tasks, goals, or interests in the formation of macrostructures" (van Dijk et Kintsch, 1983, p.192).

Les processus de compréhension et de production de texte sont nécessairement impliqués dans l'indexation en tant que principes de base permettant de rendre compte de certaines étapes du processus d'indexation. Mais les exigences propres à la tâche demandent que l'indexation soit examinée plus en fonction de l'ensemble de ses contraintes et dimensions propres si l'on veut parvenir à saisir l'importance de leur rôle dans la problématique de l'incohérence interindexeurs.

### 2.3 Indexation en tant que situation de résolution de problème

La section précédente a permis de démontrer qu'il est possible de rendre compte d'un ensemble d'activités cognitives reliées à l'indexation à l'aide des théories de la compréhension de texte, mais que la tâche d'indexation ne peut être considérée uniquement comme un processus de compréhension de texte. En effet, cette compréhension est orientée en fonction de macrostratégies externes au processus de compréhension de texte qui doivent tenir compte d'un contexte particulier qui est celui d'une tâche à effectuer, en l'occurrence, l'indexation (les indexeurs sont limités dans le temps, ils doivent respecter des normes, ils doivent tenir compte des besoins des usagers, etc.).

La nature cognitive de cette tâche et les diverses contraintes qui y sont associées permettent d'envisager l'indexation en tant que situation de résolution de problème. Ainsi, puisque l'on cherche à repérer les causes des variations interindexeurs par la mise à jour des processus intellectuels impliqués dans cette tâche, il apparaît pertinent d'analyser le travail des indexeurs à l'aide d'une autre théorie provenant des sciences cognitives, soit la résolution de problème (Holyoak, 1990; Newell et Simon, 1972).

### 2.3.1 Description de la résolution de problème

Définissons de prime abord ce qui caractérise la résolution de problème. Une activité peut être qualifiée de résolution de problème lorsqu'un individu oriente ses actions en fonction d'un but (*goal-directed*), que l'atteinte de ce but implique une séquence de processus mentaux plutôt qu'un seul, et que les processus impliqués sont nécessairement de nature cognitive (Eysenck, 1984).

En termes plus formels, la résolution de problème peut être considérée comme un processus lors duquel le sujet part d'un état de connaissance initial et se déplace à l'intérieur de l'espace-problème pour arriver à l'état final ou à la solution (Newell et Simon, 1972). L'espace-problème est la représentation que le sujet se fait de la tâche et comprend essentiellement l'ensemble des connaissances jugées potentiellement utiles pour la résolution. Ceci inclut la représentation du but et les opérateurs pouvant être utilisés. Les opérateurs comprennent les actions (cognitives ou physiques) pouvant être appliquées et les conditions nécessaires pour l'application des actions permettant le passage d'un état à un autre, et ainsi réduire la distance du but. L'espace-problème comprend également les connaissances stratégiques du sujet, c'est-à-dire des connaissances se rapportant à des méthodes de résolution acquises avec l'expérience de la résolution de problèmes appartenant à un même domaine (Greeno et Simon, 1988).

Il est admis qu'un individu ne se déplace pas au hasard dans l'espace-problème, mais qu'il procède plutôt par une recherche orientée en fonction du but, impliquant l'utilisation de plans et de sous-buts. Ainsi, lorsque le but principal ne peut être atteint directement, le problème est alors décomposé en un ensemble de sous-buts, dont l'organisation constitue la structure de buts globale du problème.

La résolution de problème est alors considérée comme un processus de recherche (*search control*) à l'intérieur de l'espace-problème qui s'articule en un cycle de trois étapes appliqué de manière récursive (Newell, 1980):

- 1) Sélectionner un état; sélectionner un opérateur.

- 2) Appliquer l'opérateur à l'état, produire un nouvel état.
- 3) Décider si le but est atteint; décider d'arrêter; décider de conserver le nouvel état.

La première étape d'une étude visant à comprendre les processus mis en œuvre lors d'une activité de résolution de problème consiste à faire une analyse de la tâche. Cette analyse fournit des informations concernant les contraintes imposées par la nature même du problème permettant de circonscrire les opérations et les états potentiels de l'espace-problème afin d'être en mesure de représenter adéquatement le comportement (physique ou cognitif) de l'individu.

Soulignons que la résolution de problème peut être caractérisée non seulement par la poursuite d'un but, mais par la recherche d'une façon optimale d'atteindre ce but, que ce soit le plus rapidement ou le plus efficacement possible. L'individu utilise alors une stratégie de résolution, appelée heuristique, qui lui permet de chercher de manière sélective parmi le nombre élevé d'alternatives, les options jugées les plus susceptibles de conduire à une solution satisfaisante, ou de répondre à un *satisficing*<sup>1</sup>. Ceci s'applique particulièrement pour les problèmes mal définis où la vérification de toutes les alternatives s'avère trop fastidieuse.

### 2.3.2 Stratégies d'indexation

Bertrand (1993) est l'une des rares auteures à avoir examiné le travail des indexeurs à l'aide de la théorie de résolution de problème. Ses recherches (Bertrand et al., 1994; Bertrand et Cellier, 1995; Bertrand et al., 1996) ont permis d'identifier les différences de stratégies d'indexation entre des groupes d'indexeurs novices et experts.

Précisons qu'un indexeur est considéré expert lorsque le domaine du document à traiter et le langage documentaire utilisé lui sont familiers, tandis qu'il est considéré novice lorsqu'il débute dans la pratique de l'indexation et par conséquent ne connaît aucun langage documentaire particulier et n'est pas familier avec le domaine des documents à indexer.

Premièrement, Bertrand (1993) identifie une procédure d'indexation commune pour l'ensemble des indexeurs. À l'étape d'exploration du document, 93 % des indexeurs parcourent le titre, le résumé et la table des matières et par conséquent, sélectionneront les notions essentielles à partir de ces sections. Tous les novices limitent leur consultation presque exclusivement à ces trois parties. Dans 65 % des cas, les indexeurs procèdent à l'extraction suivie de la transformation des termes retenus en termes d'indexation; ce qui correspond à la prescription de la norme ISO (1985). Pour l'étape de traduction, 93 % des indexeurs consultent le langage documentaire pour choisir le terme d'indexation correspondant à la notion retenue, tandis que seulement 7 % des indexeurs transforment directement une notion en terme d'indexation, sans avoir recours au langage, ce qui nécessite une connaissance approfondie du langage documentaire.

Bertrand (1993) relève également des différences entre novices et experts au niveau des stratégies procédurales. Premièrement, les stratégies d'exploration semblent être guidées en partie par la représentation de la structure conventionnelle du document. Ainsi, il apparaît que les novices ont une représentation plus rigide de la structure des documents, par exemple que la table des matières se trouvera au début du document. Les indexeurs professionnels ayant analysé un plus grand nombre de documents de structures différentes ont acquis une connaissance des différentes structures possibles des documents, ce qui rend leur représentation plus riche et oriente d'autant plus leur exploration.

Deuxièmement, au niveau de la sélection des concepts, novices et experts sont caractérisés par un objectif différent. Les novices visent à "extraire des informations principales qui soient relativement générales pour faciliter l'accès au langage documentaire" (Bertrand, 1993, p.193) et les experts visent à extraire des informations principales qui soient aussi des termes d'indexation ou, éventuellement, compatibles avec des termes d'indexation.

---

<sup>1</sup> Nous reviendrons plus loin sur la notion de *satisficing*, liée au travail de Simon (1973, 1990) en relation avec les problèmes mal définis.

En conséquence, on remarque que les stratégies de conversion en termes d'indexation sont également tributaires de la connaissance du langage documentaire. À ce sujet, Bertrand (1993) et Bertrand et al. (1994) répertorient trois types d'objectifs poursuivis lors de l'exploration du langage. Les experts peuvent être caractérisés par les deux premiers objectifs. L'objectif de la *validation d'un concept* qui consiste à vérifier si une notion est acceptée en tant que terme d'indexation. L'indexeur consulte alors le seul terme d'indexation et explore parfois les termes reliés à ce dernier. L'objectif de *traduction d'une notion ou de spécification d'un concept* qui consiste à trouver des termes d'indexation qui traduisent ou spécifient les notions sélectionnées du document. Cet objectif se distingue du précédent du fait que l'indexeur ne recherche pas un terme particulier d'indexation, mais consulte plutôt le langage afin de préciser une notion floue ou exprimer la relation entre plusieurs notions. Quant aux novices, ils sont caractérisés par l'objectif *d'extraction de nouveaux concepts-termes d'indexation* qui consiste à utiliser le langage comme aide à l'extraction des concepts du document. Les novices n'ayant qu'une connaissance minimale de l'indexation et n'étant pas familiers avec le langage documentaire se servent du langage pour repérer les notions importantes du document et ainsi se construire une représentation conceptuelle de l'ensemble des idées exprimées par le document. Ils choisiront alors un terme d'entrée générique dans le langage afin d'avoir accès à un vaste ensemble de termes d'indexation. Concrètement, ceci leur permet de diminuer le temps consacré à la recherche des notions principales ainsi que les va-et-vient dans le langage (Bertrand, 1993). Ainsi, ces résultats ont permis de révéler une fonction du langage documentaire qui n'est pas répertoriée par les normes d'indexation : "Le langage est aussi utilisé comme outil d'aide à la construction de ces représentations (représentations des documents)" (Bertrand et al., 1994, p.43). Le langage tient non seulement un rôle majeur à l'étape de conversion des concepts en termes d'indexation, mais parfois à l'étape d'extraction des concepts. Notons que chacun de ces objectifs n'est pas exclusif à un seul groupe d'indexeurs, mais caractérise plus particulièrement les stratégies adoptées par l'ensemble des indexeurs de ce groupe.

Principalement, les recherches de Bertrand (1993) et Bertrand et al., (1994) ont permis d'identifier les stratégies générales d'indexation propres aux indexeurs novices et experts, lesquelles se distinguent par la nature des connaissances utilisées, ainsi que le type d'objectifs poursuivis. La stratégie-novice consiste en une *indexation assistée par l'outil*

*documentaire* comparativement à celle des experts qui est une *indexation orientée par les connaissances sur des termes d'indexation potentiels*.

Ainsi, la stratégie d'indexation des novices, *indexation assistée par l'outil documentaire*, est guidée par la recherche de termes relativement généraux à l'étape d'extraction pour faciliter le repérage de termes d'indexation correspondants, c'est-à-dire qui constituent une entrée dans le langage documentaire. Leur schéma référentiel étant pauvre, les novices procèdent par traitements ascendants, utilisant les indices de surface du document (superstructure) et du langage documentaire (réseau de termes d'indexation) pour construire une représentation du contenu du document qui soit constituée directement de termes d'indexation. À ces traitements ascendants est juxtaposée une exploration relativement succincte du document, limitée aux "titre, résumé et grands titres de la table des matières, si ces derniers sont peu nombreux" (Bertrand, 1993, p.244).

Ces observations permettent, d'une part, de conforter le constat qu'une macrostructure peut être construite uniquement à partir d'indices de surface (Rossi et Bert-Erboul, 1991 in Bertrand, 1993). Cependant, le recours aux indices de surface ne peut pallier le manque de connaissances nécessaires à une compréhension adéquate du texte et du langage documentaire permettant de construire une macrostructure du document qui soit précise et cohérente. En conséquence, les termes d'indexation choisis par les novices demeurent généraux et leur pertinence peut être questionnée puisqu'il s'agit d'une évaluation basée sur le réseau des termes d'indexation et non sur le contenu du document (Bertrand, 1993).

D'autre part, cette stratégie fait ressortir le caractère dirigé de l'activité de l'indexeur en fonction de la contrainte première de l'indexation, soit l'utilisation d'un langage documentaire pour la production d'une liste de termes d'indexation. On remarque ainsi que malgré des connaissances parcellaires du problème, le novice utilise les éléments à sa disposition (en l'occurrence les éléments de surface) de manière optimale pour répondre aux exigences de la tâche.

La stratégie d'indexation générale adoptée par les experts, *indexation orientée par les connaissances sur des termes d'indexation potentiels*, témoigne également d'une utilisation

optimale des connaissances de la tâche. Ainsi, leur connaissance du langage documentaire est utilisée dès la phase d'exploration du document, leur permettant de reconnaître et d'extraire les notions qui peuvent être traduites directement en termes d'indexation. L'évaluation de la pertinence d'une notion repose en grande partie sur ce critère. D'ailleurs, les résultats de Bertrand et al. (1994) démontrent une augmentation significative de la cohérence intra-indexeur (recouvrement entre concepts sélectionnés à l'étape d'extraction et termes d'indexation) avec des taux moyens de 64 % pour les indexeurs connaissant le langage et de 53 % pour les indexeurs n'étant pas familiers au langage.

Pour les experts, la nature des consultations du langage se limitent à une vérification ou une précision des termes d'indexation repérés dès la première étape de l'indexation. Toutefois, Bertrand (1993) a observé quelques "recherches vaines" chez ces indexeurs, qui se retrouvent par contre en plus grand nombre chez les novices.

Cette stratégie, associée à des traitements plus descendants, permet à l'expert d'adapter son niveau de description du contenu du document à celui du langage documentaire. Ainsi, bien que sa structure conceptuelle du domaine lui permette de construire une macrostructure élaborée et de repérer des notions exprimées de manière implicite, l'indexeur évaluera la pertinence des notions implicites qui correspondent à un terme d'indexation.

De plus, il semble que ces indexeurs évaluent la représentativité d'un terme d'indexation "en comparant la situation évoquée dans l'ouvrage aux situations habituellement représentées par ces termes d'indexation" (Bertrand, 1993, p.244). En d'autres mots, l'indexeur fait référence à sa connaissance du fonds documentaire pour attribuer des termes d'indexation qui situent le document par rapport à l'ensemble des documents contenus dans le système.

Il ressort de ces observations que les indexeurs experts ont effectivement une représentation plus riche du problème, ce qui leur permet d'évaluer un élément de solution ou la solution en fonction de plusieurs dimensions simultanément. Ceci rejoint une recommandation de la norme ISO (1985) qui souligne que l'indexation gagnera en qualité si

l'indexeur possède "une bonne connaissance du domaine couvert par les documents qu'il indexe. Il doit comprendre les termes rencontrés dans les documents aussi bien que les règles et les procédures du langage d'indexation spécifique" (p.4). Il n'est donc pas surprenant de constater que le taux de cohérence interindexeurs le plus élevé (40 %) obtenu par Bertrand (1993) appartient au groupe d'indexeurs experts. Quant aux indexeurs novices, ils atteignent une cohérence de 31 %.

La stratégie d'indexation propre aux indexeurs experts identifiée par Bertrand (1993) corrobore l'assertion théorique de Beghtol (1986) qui soutient qu'un indexeur ne pourra faire une analyse impartiale d'un document puisqu'il sait que les notions extraites doivent être transposées en termes d'indexation. Tout comme l'indexeur novice, mais à un niveau différent, l'indexeur expert trahit l'objectivité de la représentation du contenu du document en orientant son analyse en fonction des contraintes du langage documentaire. L'introduction de biais dans leurs analyses s'explique par le but commun de réaliser la tâche de manière efficace en mettant à profit l'ensemble de leurs connaissances.

Malgré une stratégie d'indexation et une structure de but similaires, le taux de cohérence interindexeurs des experts n'excède pas 40 %. Bien qu'il soit légitime de s'interroger sur les raisons d'un tel degré d'accord, les analyses de Bertrand (1993) ne permettent pas d'expliquer pourquoi des indexeurs de profil identique diffèrent de manière relativement marquée dans leurs choix de termes d'indexation.

Il semblerait donc, que les composantes de l'indexation les plus susceptibles de fournir des explications concernant les variations du choix de termes d'indexation, ne peuvent être mises de l'avant par la seule comparaison des procédures d'indexation ou des structures de buts. Par exemple, les "recherches vaines" effectuées par les indexeurs experts dans le thésaurus identifiées par Bertrand (1993) constituent un indice que leurs processus cognitifs doivent être examinés sous un angle différent. En effet, il serait intéressant de comprendre les raisons qui motivent un indexeur expert à faire une fouille élaborée du langage et ce, malgré sa connaissance des possibilités offertes par le langage. À notre avis, ces fouilles ne doivent pas être attribuées exclusivement à un manque de connaissance d'une portion du langage, mais témoignent probablement de la prise en

compte d'une dimension du problème autre que le langage documentaire pour décider de la pertinence d'un terme.

À notre avis, si l'on veut pouvoir rendre compte des différences interindexeurs, il ne faut pas limiter l'analyse de la tâche d'indexation à la structure générale de résolution, mais considérer la nature même du problème, c'est-à-dire examiner d'autres éléments qui caractérisent le problème et qui en font un problème bien ou mal défini.

### 2.3.3 Problèmes bien et mal définis

Les recherches à la base de la théorie de résolution de problèmes se sont surtout intéressées à des problèmes relativement bien définis (*well-defined*) tels les problèmes de logique formelle (par exemple la tour de Hanoï, les énigmes, etc.), les problèmes de cryptarithmétique (Newell et Simon, 1972), les problèmes d'algèbre (Wickelgreen, 1974), etc. Par exemple, examinons un problème simple de logique formelle comme la tour de Hanoï (Anzai et Simon, 1979) :

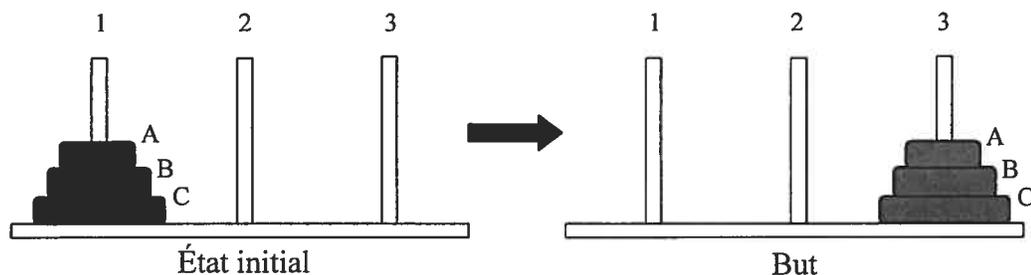


Figure 2  
Tour de Hanoï.

Le sujet se trouve devant trois tiges dont la première est occupée par trois cercles, ou disques, de grosseur décroissante empilés l'un sur l'autre. Il doit reconstituer ce pylône de trois cercles en déplaçant les cercles un à un de leur tige initiale vers la troisième tige, en exécutant le moins de mouvements et ce, sans qu'un petit cercle ne se retrouve sous un plus grand. Les contraintes de ne bouger qu'un cercle à la fois et qu'un petit cercle ne se retrouve sous un plus grand limitent les déplacements, et en conséquence le nombre d'états potentiels du problème. De même, le point de départ est très bien défini et l'état final unique permet

une évaluation précise de la réussite ou de l'échec de la résolution du problème. Tout comme les problèmes d'algèbre et de physique, plus complexes mais possédant néanmoins une solution unique, ce qui caractérise ce type de problème est que la presque totalité de l'information nécessaire à la résolution est contenue dans l'énoncé même du problème. La procédure de résolution suivie est souvent prédéfinie et requiert l'application de règles strictes, voire d'algorithmes. En fait, l'analyse de la tâche de ces problèmes permet de définir *a priori* presque toutes les opérations, les états possibles et les règles gouvernant les différentes alternatives.

Mais ce ne sont pas tous les types de problèmes qui permettent une description et une organisation précise de l'ensemble de leurs composantes. La majorité des problèmes rencontrés dans le quotidien sont de nature beaucoup moins structurée et définie. Simon (1985) propose trois caractéristiques permettant la distinction entre problèmes bien et mal définis (*ill-defined*), en soulignant que cette distinction n'est pas catégorique mais qu'il s'agit plutôt d'un continuum partant de problèmes comme la tour de Hanoï à des tâches telles que l'achat d'une maison ou la composition d'une symphonie<sup>2</sup>. Premièrement, le critère permettant de déterminer si le but est atteint est plus complexe et moins défini. Deuxièmement, l'information nécessaire à la résolution n'est pas contenue entièrement dans l'énoncé du problème. De fait, l'information nécessaire peut provenir de références extérieures au problème, la délimitation de l'information pertinente étant elle-même très vague. Et finalement, il n'existe pas d'opération "légale" prédéfinie afin de connaître l'ensemble des alternatives à chaque étape (Simon, 1985). En fait, un problème peut être mal défini en fonction de son état initial, intermédiaire ou final, et il se peut que la procédure de résolution comporte des zones floues et doive être élaborée par le sujet.

Par ailleurs, Reimann et Chi (1989) soulignent que la nature du problème peut nécessiter un changement d'approche pour comprendre la résolution d'un problème. L'accent peut être mis sur le type de connaissances qu'un individu utilise pour se construire une représentation adéquate du problème menant à une solution, plutôt qu'uniquement sur le type de stratégies qui guide sa résolution.

---

<sup>2</sup> Pour les besoins de notre recherche nous ne distinguerons ici que les problèmes bien et mal définis. On trouvera une typologie des problèmes chez Chandrasekaran (1983) et Greeno et Simon (1988).

L'utilisation de ces connaissances est également tributaire de l'expertise du sujet face au problème à résoudre. Premièrement, les experts possèdent une plus grande expérience du domaine de la tâche et peuvent en conséquence utiliser leur connaissance de manière plus efficace que les novices (Reimann et Chi, 1989). Il a été observé qu'un novice et un expert possédant les mêmes connaissances déclaratives différencieront dans leur utilisation de ces connaissances. En effet, comparativement à la connaissance plus théorique du novice, la connaissance "pratique" de l'expert lui permet de se construire une meilleure représentation de l'utilisation potentielle des différents éléments de connaissance. Entre autres, ceci permet à l'expert de réduire considérablement le nombre d'alternatives à considérer lors de sa fouille. L'expertise peut donc être caractérisée en tant que raffinement continu de la connaissance d'un domaine par l'application répétée de ce savoir (Reimann et Chi, 1989). Non seulement l'expert "connaît" plus que le novice, mais il existe également des différences qualitatives entre novices et experts dans la façon d'organiser et d'utiliser les connaissances pour la résolution. Ainsi, il semble que le novice organise ses connaissances en fonction des éléments de surface du problème, tandis que l'expert semble avoir une organisation basée sur les notions plus profondes ou abstraites concernant le problème (Elio et Scharf, 1990).

Dans le contexte des problèmes mal définis, on doit revenir sur une notion fondamentale relative à la résolution de ce type de problèmes, soit la notion de *satisficing* développée par Simon (1955, 1956 in Agosto, 2002). Ce concept est issu des recherches sur la prise de décision.

Simon (1955, 1956 in Agosto, 2002) suppose que le processus de prise de décision d'un individu prend place à l'intérieur de limites temporelles et cognitives ne lui permettant pas d'évaluer le résultat de toutes les options avant de prendre une décision. Ainsi, l'humain agirait de manière rationnelle, mais à l'intérieur de limites pragmatiques, choisissant la meilleure option selon sa connaissance de la situation.

En d'autres termes, il s'agit de choisir une solution satisfaisante et suffisante qui correspond aux objectifs du "preneur de décision", mais qui n'est pas nécessairement la

solution optimale: "Satisficing involves setting an acceptable level or aspiration level as a final criterion and simply taking the first acceptable option" (Newell et Simon, 1972, p.681). La notion de *satisficing* implique donc qu'un individu détermine les conditions auxquelles doit répondre une bonne solution et cesse la recherche dès qu'une solution répond de manière raisonnable à ses critères (Simon, 1990).

Ajoutons que le *satisficing* ne limite pas l'individu à un seul facteur de décision ou critère d'arrêt:

When the criterion of problem solution or action has more than one dimension, there is the matter of calculating the relative merits of several alternatives, one of which may be preferred along one dimension, another along another... The satisficing rule... stipulates that the search stops when a solution has been found that is good enough along all dimensions. (Simon, 1979, in Agosto 2002, p.17)

D'ailleurs, l'individu ajuste et réévalue les options en fonction de l'état du problème et la nature des contraintes qui prennent forme tout au long du processus de résolution. Ceci implique nécessairement des compromis, l'individu devant alors s'en tenir à un *satisficing* qui regroupe un ensemble plus ou moins important des conditions visées.

Dans le contexte d'une indexation, on peut supposer que l'indexeur élabore une stratégie de résolution qui lui permette d'atteindre une solution répondant aux conditions du *satisficing* visé, c'est-à-dire que l'indexeur juge sa liste de descripteurs complétée lorsqu'elle répond de manière raisonnable aux exigences de la tâche.

### 2.3.4 Indexation comme problème mal défini

En ce sens, l'indexation peut être considérée comme un problème mal défini puisque c'est l'indexeur qui détermine que la solution est atteinte, donc que le problème est résolu. En d'autres termes, l'indexeur ne dispose pas de réponse toute préparée, mais doit construire celle-ci à partir d'une analyse de la situation et des connaissances qu'il possède, faisant passer un objet (matériel ou conceptuel) d'un état initial à un état souhaité par l'application d'opérateurs (Leplat, 1980). La sélection et l'application d'opérateurs n'est pas aléatoire,

mais dépend bien d'un contrôle stratégique effectué par le sujet en fonction de ses connaissances et de son évaluation de l'état atteint.

Une brève analyse de l'indexation permet de faire ressortir les éléments de la tâche qui sont relativement mal définis, c'est-à-dire ceux pour lesquels l'indexeur pourra établir sa propre série de critères d'application.

De prime abord, la structure des procédures de l'indexation semble relativement bien définie par la norme internationale ISO (1985). L'indexeur sait que le parcours des parties les plus informatives du document lui permet d'identifier les thèmes et de sélectionner les notions essentielles. *A priori*, la procédure de traduction apparaît également bien circonscrite. La figure 3 présente l'organigramme du processus d'indexation (ISO, 1985).

D'ores et déjà, l'importance accordée à la description de l'étape de traduction comparativement à l'étape de l'identification des notions essentielles permet de pointer sur une des principales zones mal définies dans le processus. Comme il a été souligné précédemment, rien dans la norme n'indique de façon précise la manière dont l'indexeur doit s'y prendre pour évaluer l'*indexabilité* d'une notion en fonction de l'ensemble des critères devant être respectés (critères de qualité, besoins des usagers, etc.), les stratégies cognitives impliquées lors de l'analyse n'étant pas abordées.

De même, plusieurs cases de l'organigramme concernant l'étape de traduction ne peuvent être définies qu'en référence directe à l'activité intellectuelle de l'indexeur. Entre autres, les cases "considérer le descripteur" et "considérer chacun des termes" renvoient aux critères d'évaluation qui seront utilisés par l'indexeur pour décider du choix ou du rejet d'un terme. Il apparaît de façon évidente ici, que l'indexeur doit utiliser son propre jugement, ou sa connaissance de la tâche, pour prendre des décisions éclairées.

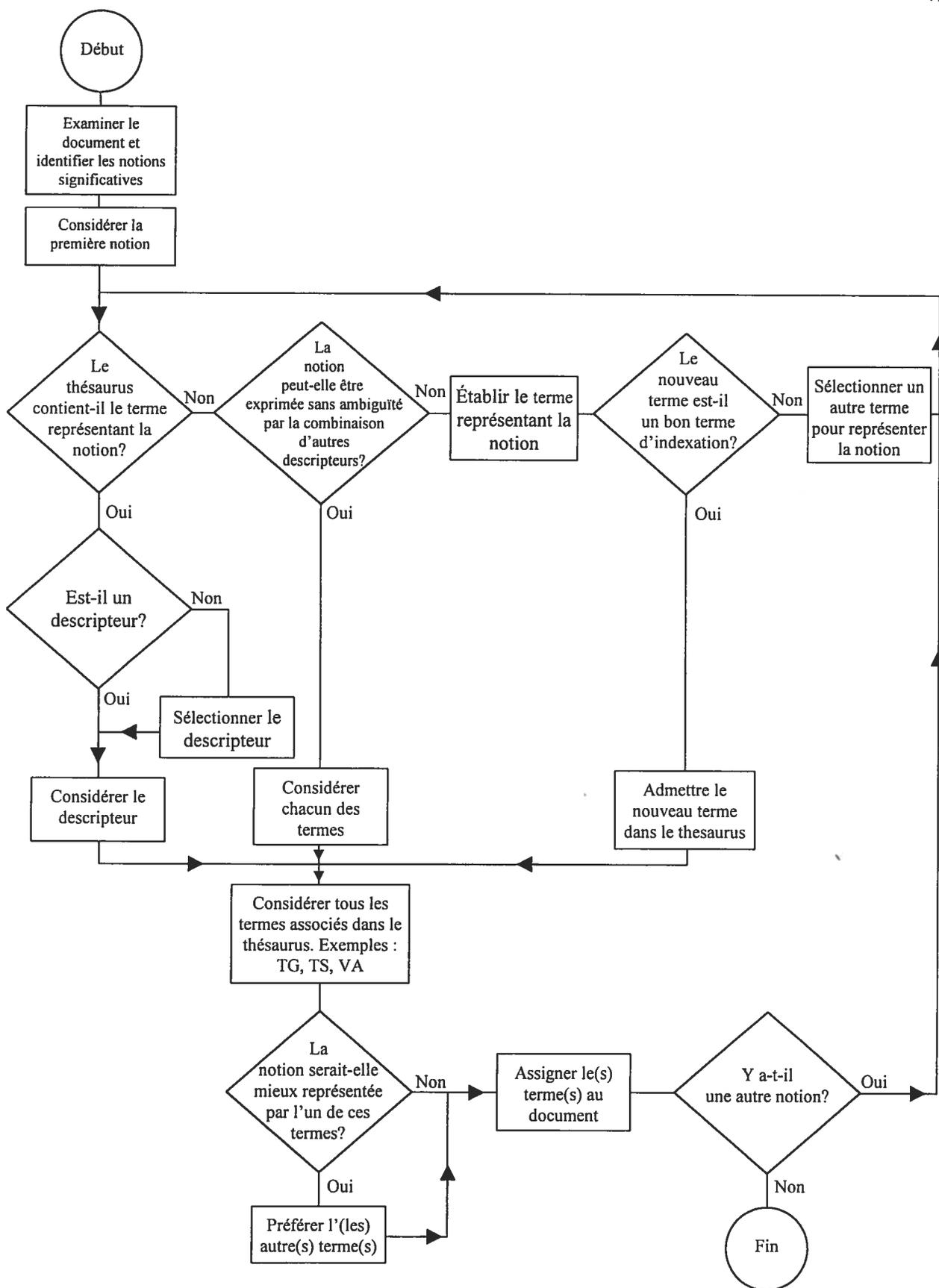


Figure 3  
Organigramme du processus d'indexation utilisant un thésaurus.

Ce type de représentation sert à décrire la procédure chronologique de l'indexation, mais demeure cependant incomplète. Il ne s'agit que d'une description de l'ensemble général des actions, pouvant à la rigueur être identifiées par la seule observation des gestes. En effet, cette description ne permet pas de saisir la prise en compte simultanée des différentes étapes d'indexation ou l'influence qu'exercent les différentes contraintes (temps de production, critères de qualité, besoins des usagers, etc.) sur le choix de termes.

Dans cette perspective, il serait intéressant d'identifier les éléments de connaissance reliés à l'indexation qui sont utilisés par un indexeur pour évaluer que la solution (sa liste de descripteurs) répond aux exigences de la tâche.

### Hypothèse de travail

Avant de poser l'hypothèse générale qui va guider nos analyses des processus intellectuels des indexeurs, rappelons brièvement notre démarche. Nous avons tenté de démontrer que les études sur la problématique de la cohérence interindexeurs, qui se sont attardées à l'identification de variables liées au seul produit, ne constituaient pas une approche appropriée pour rendre compte des variations interindexeurs. Il a donc été établi que pour mieux comprendre les causes de l'incohérence, il fallait plutôt s'attarder aux dimensions cognitives impliquées dans la tâche d'indexation. À cet effet, nous avons présenté les résultats de plusieurs recherches ayant examiné l'indexation en tant que processus de compréhension de texte. Cependant, nous avons démontré que le processus d'indexation s'articule dans un contexte qui n'est pas uniquement de la compréhension de texte, mais qui est plutôt orienté en fonction d'une tâche particulière comportant ses propres contraintes. Il semblait donc pertinent d'examiner l'indexation à l'aide de la théorie de résolution de problème et plus particulièrement en tant que problème mal défini, l'indexation correspondant à ce type de problème.

En conséquence, nous supposons que l'incohérence relative entre indexeurs experts prend sa source dans la représentation personnelle de chaque indexeur en ce qui concerne la

prise en compte des composantes mal définies du problème de l'indexation. Ceci reprend, de manière plus spécifique, l'argument de Milstead (1994) précédemment évoqué:

In fact, it seems reasonable to assume that once differences in training, subject expertise, and familiarity with the vocabulary and indexing policies are accounted for, what remains are differences in cognitive experiences with the document. (p.578)

Notre objectif consiste à expliquer la problématique de la cohérence interindexeurs sous un angle différent en prenant comme base l'hypothèse générale suivante: Avec l'expérience, l'indexeur expert a assimilé les normes et règles de l'indexation de manière personnelle, et en conséquence a développé sa propre méthode (critères privilégiés et stratégie personnelle) pour produire une indexation satisfaisante.

En d'autres termes, qu'est-ce qui permet à l'indexeur de déterminer que son indexation est complétée? Puisqu'il doit se fier à son propre jugement, sur quoi base-t-il l'évaluation de ses choix de termes? Qu'est-ce qu'un terme *indexable* pour l'indexeur expert? Est-ce que des indexeurs experts de profil similaire utilisent les mêmes critères? Si oui, les pondèrent-ils de la même manière?

Ainsi, nous précisons l'hypothèse générale en supposant que l'examen détaillé des divers éléments de connaissance utilisés par les indexeurs experts permettra d'avoir un portrait précis de ce que doit comporter une "bonne indexation" (Bell et Jones, 1979) ou une solution répondant à un *satisficing* personnel à chaque indexeur. En retour, la systématisation de ces éléments de connaissance ou critères devrait permettre de fournir une nouvelle explication à la variation interindexeurs de choix de termes.

### *2.3.5 Application des éléments de connaissance*

Avant de procéder à une expérimentation pour tester notre hypothèse, il est opportun d'examiner les écrits en SI pour voir dans quelle mesure on y trouve des indices sur la façon dont ces divers éléments peuvent être pris en considération par l'indexeur.

Les éléments auxquels nous nous attarderons se subdivisent en deux groupes: Les critères relatifs à la couverture sémantique du contenu du document, c'est-à-dire ceux qui font référence plus directement à l'adéquation entre les termes choisis et le contenu des documents, soit les critères de qualité de l'indexation; et les éléments relatifs au contexte de la tâche, c'est-à-dire les éléments qui font référence à l'adéquation entre les termes choisis et l'utilisation qui en sera faite, soit la prise en compte de l'utilisateur et la pertinence.

En ce qui concerne les critères de qualité de l'indexation, nous limiterons la révision aux critères mentionnés dans la norme ISO (1985) puisqu'elle décrit les principes de base de l'indexation et qu'en conséquence tous les indexeurs s'y sont référés lors de leur formation. De plus, ces critères représentent les aspects de la tâche qui, en théorie, doivent nécessairement être respectés par l'indexeur lors de son analyse.

Nous examinerons donc plus en détails les critères de qualité de l'indexation, notamment, l'exhaustivité, la spécificité et l'exactitude. Rappelons que la norme souligne que "le critère principal, dans la sélection des notions, doit toujours être la valeur potentielle d'une notion en tant qu'élément dans l'expression du contenu d'un document et dans sa recherche documentaire" (ISO, 1985, p.3) et insiste sur la prise en compte des besoins des usagers afin de "déterminer si certains termes ou descripteurs sont susceptibles de produire des combinaisons fausses et de créer des sorties non pertinentes" (ISO, 1985, p.3). Nous passerons ensuite en revue l'évaluation de la valeur potentielle d'une notion que nous associons au critère de pertinence ou d'*indexabilité* ainsi que la dimension besoins des usagers qui ne sont définis qu'en termes vagues par la norme.

Précisons que les normes ISO (1985), BSI (1984) et AFNOR (1978) insistent principalement sur les qualités d'exhaustivité et de spécificité et que les politiques d'ERIC (Educational Resources Information Center, 1980) comportent une section spéciale sur la spécificité. De même, les études s'intéressent principalement aux critères d'exhaustivité, de spécificité et de pertinence. En conséquence, nous traiterons de ces critères de manière plus élaborée.

## Exhaustivité

L'exhaustivité est généralement associée au nombre de notions retenues pour représenter le contenu intégral de l'ouvrage (ISO, 1985). Il est établi qu'une indexation exhaustive ou en profondeur comportera une douzaine de termes par opposition à une indexation superficielle qui ne retient que les notions principales, soit de trois à cinq termes (Bertrand-Gastaldy, 1990; ERIC, 1980). Selon Sparck Jones (1973), les différents niveaux d'indexation sont associés à des sources différentes: Les termes d'une indexation superficielle sont tirés du titre, et ceux d'une indexation en profondeur sont extraits du résumé. Ainsi, il est possible en principe, de reconnaître l'application de l'exhaustivité chez un indexeur à son nombre de notions retenues.

Mais au-delà du nombre de termes, de quelle manière la prise en compte de cette qualité influence-t-elle les décisions de l'indexeur?

À ce sujet, la norme ISO (1985) souligne qu'un examen minutieux des sections les plus informatives du document devrait permettre à l'indexeur d'identifier toutes les notions traitées dans le document et recommande de retenir tous les termes qui ont une valeur potentielle pour les usagers. Aussi, l'avènement des réseaux d'information demande qu'une attention particulière soit apportée aux différents aspects ou facettes pouvant intéresser des usagers provenant de disciplines autres que celles couvertes par le système.

D'ailleurs, plusieurs auteurs s'intéressent aux effets de la variation de l'exhaustivité sur la performance du repérage (Soergel, 1994; Sparck Jones, 1973) visant à mettre au point des techniques permettant de repérer des références pertinentes indépendamment du niveau de profondeur de l'indexation (Burgin, 1991; Salton et Buckley, 1990). La perspective adoptée pour étudier l'exhaustivité est celle du repérage plutôt que celle de l'indexeur appliquant l'exhaustivité, et l'on évalue la qualité des termes retenus en fonction de la satisfaction de l'utilisateur.

En rapport à la recherche documentaire, il faut savoir que plus l'exhaustivité est grande, plus le taux de rappel est élevé, donc, le bruit risque d'augmenter au repérage

provoquant ainsi une baisse du taux de précision. De fausses liaisons entre des termes trop nombreux signalent alors à l'utilisateur des documents qui ne traitent que très superficiellement des termes qu'il recherche. Inversement, une indexation plus superficielle permet d'augmenter le taux de précision mais accroît le silence, c'est-à-dire le risque de ne pas fournir à l'utilisateur une information qui lui serait utile. En effet, si les concepts sont absents de l'indexation, l'utilisateur ne peut repérer les documents qui en traitent. Le manque d'exhaustivité provient d'une indexation qui ne contient pas suffisamment de termes pour décrire de façon précise le contenu du document. Par conséquent, l'utilisateur doit faire une consultation élaborée du document pour savoir sous quel angle le concept est abordé. D'une certaine manière, il doit terminer le travail de l'indexeur.

L'exhaustivité d'une indexation ne doit pas être évaluée seulement en fonction du nombre de termes d'indexation attribués, mais plutôt, en se basant sur la proportion de l'information indexée versus la quantité totale d'information contenue dans le document. Pour un document de quelques pages par exemple, cinq termes constituent une indexation exhaustive alors qu'elle sera superficielle pour un rapport de recherche. D'ailleurs, Hutchins (1975) ajoute que l'exhaustivité "is a notion definable only in relation to particular documents and to particular information systems" (p.111). En effet, le niveau d'exhaustivité est également tributaire des possibilités offertes par le langage documentaire utilisé. À l'étape d'extraction, un indexeur peut faire un relevé exhaustif des notions traitées dans le document et se voir contraint de réduire cette représentation si le langage documentaire ne permet pas ce niveau de description; l'application d'une grande exhaustivité nécessite donc un langage documentaire spécifique.

Les indices avancés dans la littérature sur ce qui doit être inclus dans une indexation pour respecter le critère d'exhaustivité peuvent être résumés en ces quelques recommandations :

- Identifier toutes les notions qui ont une valeur potentielle pour les utilisateurs d'un système d'information dans un document (ISO, 1985).
- Indexer tous les concepts significatifs traités par le document (ERIC, 1980).

- Présenter de manière exhaustive les concepts importants d'un document (Bonville 1989, in Patch-Neveu, 1989).
- Tous les concepts utiles doivent être retenus (Chaumier, 1988).
- L'exhaustivité d'une indexation renvoie à la quantité d'information moins importante qui est retenue (Hutchins, 1975).
- Lesquels des concepts associés à cette entité (document) sont suffisamment importants pour en justifier l'indexation? (Soergel, 1994).

Il ne semble donc pas exister de règles précises pour déterminer l'exhaustivité, mais seulement de vagues recommandations (Bertrand-Gastaldy, 1990).

Pour certains auteurs cependant (van Slype, 1987 in Bertrand-Gastaldy, 1990; Chaumier 1988), l'indexeur devrait appliquer la règle d'exhaustivité au moment du choix des concepts à représenter et appliquer la règle de spécificité à l'étape de traduction. Ainsi, l'indexeur tirerait les termes d'indexation les plus spécifiques à partir d'une représentation exacte et complète du contenu du document. Il serait à même de faire des compromis éclairés en fonction des contraintes du langage documentaire. En effet, l'application de l'exhaustivité ne signifie pas que le niveau de spécificité des termes choisis équivaut nécessairement à celui des notions du document ou à celui du vocabulaire du langage.

Il semble donc que l'exhaustivité ne peut pas être étudiée de manière isolée en fonction du contenu du document mais doit plutôt être considérée en relation avec les besoins des usagers, la spécificité, l'étendue et la spécificité du langage, et les politiques d'indexation propres à un centre de documentation. En d'autres termes, l'exhaustivité n'est pas le seul facteur déterminant le nombre de termes assignés à un document (Soergel, 1994).

Ainsi, les décisions d'un indexeur visant à produire une indexation exhaustive seront vraisemblablement influencées par l'un ou l'autre de ces facteurs puisqu'il semble peu probable qu'un indexeur expert juge son indexation satisfaisante au seul nombre de termes contenus dans sa liste finale. Évidemment, le facteur principal déterminant le niveau d'exhaustivité d'un indexeur provient des politiques d'indexation auxquelles l'indexeur doit

se conformer. Même si ces politiques privilégient une indexation d'un certain niveau de profondeur, la quantité de termes d'indexation attribués est toujours tributaire des dimensions privilégiées par le centre. Prenons, par exemple, la politique d'ERIC (1980) :

At least one major descriptor must be assigned to a document; not more than five (5) major descriptors are permitted for each document.... There is, however, no upper limit on the total number of descriptors that can be used to index a given document; the average is around eleven (11) per document. (p.414)

Il est donc plus que probable que des indexeurs d'égale expertise produisent des listes de termes d'indexation de longueur différente. Mais comme l'application de l'exhaustivité semble être liée à un ensemble de facteurs, il serait intéressant de pouvoir identifier sur quels critères repose la prise en compte de l'exhaustivité chez différents indexeurs afin d'expliquer une cause probable du manque de cohérence interindexeurs.

Concernant le processus global de l'indexation, Chaumier (1980) fait remarquer que tout indexeur peut être caractérisé en tant qu'indexeur court ou indexeur long; les deux comportements étant mutuellement exclusifs. En d'autres termes, l'indexeur aura une tendance naturelle à retenir un nombre de termes d'indexation se situant au-dessous ou au-dessus de la moyenne du nombre de termes assignés à un document pour un système particulier. L'indexeur devrait prendre conscience de cette tendance et chercher à équilibrer son indexation.

Pour les politiques d'ERIC (1980), ces tendances sont associées aux défauts les plus répandus dans la pratique de l'indexation: La sous-indexation (*under-indexing*) et la surindexation (*over-indexing*). Un document sous-indexé est représenté, soit par quelques termes génériques derrière lesquels le contenu réel du document disparaît, soit par quelques termes spécifiques tirés du titre ou du résumé (ERIC, 1980). Ce type d'indexation proviendrait de l'omission de l'indexeur à se mettre à la place des usagers. Parallèlement, une surindexation serait le résultat d'une analyse hâtive. Il est plus aisé de noter chaque notion rencontrée que de procéder à une discrimination méticuleuse de l'information réellement *indexable*. Par exemple, il peut s'agir d'un déséquilibre entre le volume du document (peu de pages) et un nombre élevé de termes d'indexation. En procédant de cette

manière, l'indexeur ne questionne pas la pertinence d'un terme en fonction de l'ampleur du traitement ou en rapport à la satisfaction de l'utilisateur au repérage.

### Spécificité (ou précision)

Rappelons que "la précision est liée à l'exactitude avec laquelle une notion particulière qui apparaît dans un document est représentée par un terme d'indexation" (ISO, 1985, p.3). Ainsi, ce critère est étroitement lié à l'exactitude puisqu'on entend par spécificité la capacité d'un terme à "réduire aux seuls documents très pertinents les documents repérés pour répondre à une question" (Bonville, 1989, in Patch-Neveu, 1989, p.26).

L'indexeur visant la spécificité doit respecter le niveau de traitement du concept choisi par l'auteur et éviter d'indexer à un niveau plus générique ou plus spécifique que celui présenté dans le document. Toutefois la norme ISO (1985) souligne deux exceptions à cette règle:

Des notions plus générales peuvent être préférées,... quand l'indexeur considère qu'un excès de spécificité pourrait agir d'une façon négative sur le fonctionnement du système d'indexation.... Selon le poids attaché à une notion par l'auteur. Si l'indexeur considère qu'une idée n'est pas complètement développée, ou qu'une allusion y est faite occasionnellement par l'auteur, l'indexation à un niveau plus général peut être justifiée. (p.3)

De même, les politiques d'ERIC (1980) indiquent certaines situations où l'indexeur doit porter une attention particulière pour le choix du niveau d'indexation. Premièrement, lorsqu'un document traite d'un concept au niveau générique et spécifique, la politique recommande de retenir les descripteurs des deux niveaux si le document contient suffisamment d'information s'y rattachant. Deuxièmement, l'indexeur pourra retenir le terme générique lorsqu'il juge que la façon dont un concept spécifique est traité dans un document ajoute de l'information pertinente par rapport au concept générique. Et troisièmement, lorsque plusieurs aspects spécifiques d'un même concept générique sont présents dans le document mais ne sont pas traités suffisamment en détail pour justifier

l'indexation de chaque concept, "the indexer must develop a *feel* for what is sufficient or reasonable information about a concept that makes it *indexable*" (ERIC, 1980, p.455).

L'application de la spécificité doit s'orienter dans la perspective de faciliter le repérage. En effet, le taux de précision au repérage dépend de la spécificité. D'ailleurs, une trop grande spécificité à l'indexation demande un effort supplémentaire à l'utilisateur lors de la formulation de sa requête. Pour une recherche globale, qui n'aurait nécessité que deux ou trois termes, l'utilisateur devra alors questionner le système sous tous les termes spécifiques appartenant au concept générique désiré pour retrouver les documents pertinents (Soergel, 1994). Donc, pour répondre simultanément à des demandes spécifiques et génériques, l'on mise sur les relations établies dans le langage documentaire (par exemple, les relations hiérarchiques des thésaurus) afin de guider l'utilisateur vers des termes plus larges ou plus précis.

Il ne faut pas confondre la spécificité du langage documentaire et celle de l'indexation. En effet, le jugement de l'indexeur n'est pas seul en cause dans le choix de termes d'indexation judicieux. L'indexeur peut identifier les concepts spécifiques et disposer d'un vocabulaire inadéquat pour les traduire, devant alors recourir à des termes plus génériques. À ce moment, tout effort intellectuel réalisé pour accéder à une grande précision est perdu.

Ces recommandations et politiques sont relativement générales, nécessitant une grande part du jugement de l'indexeur pour l'application d'un niveau de spécificité adéquat. Certains auteurs, dont Lancaster (1970, in Bertrand-Gastaldy, 1986), considèrent toutefois qu'il s'agit d'une erreur d'indexer un concept spécifique sous un terme générique. Les travaux de Svenonius (1971, in Bertrand-Gastaldy, 1986) démontrent que la qualité des termes choisis permet un repérage plus efficace qu'une quantité élevée de termes pris au hasard : " L'indexation avec 7 termes du titre donne de meilleurs résultats qu'avec 25 termes pris au hasard" (Bertrand-Gastaldy, 1986, p.10). Doit-on supposer qu'une indexation spécifique serait préférable à une indexation exhaustive? En fait, la spécificité devrait toujours dépendre de l'étendue du fonds documentaire. Ainsi, l'indexeur procéderait à une

analyse moins spécifique lorsque le fonds contient peu de documents sur un sujet et indexerait de manière plus spécifique lorsque ce nombre est élevé.

En résumé, il semble que l'indexeur peut varier le degré de spécificité de son indexation indépendamment d'un niveau de spécificité élevé du langage documentaire. Tout comme pour l'exhaustivité, la prise en compte des dimensions telles que son utilité pour les usagers, le fonds documentaire, les politiques du centre, etc., participe à déterminer la nature de la spécificité utilisée par un indexeur donné. On peut donc supposer que différents indexeurs n'auront pas la même utilisation et/ou acception de la spécificité. En conséquence, il semble plausible que la mise à jour de prises en compte différentes de la spécificité contribue à l'explication du manque de cohérence interindexeurs.

#### Exactitude (ou véracité)

L'exactitude n'est pas traitée explicitement dans la norme ISO (1985) qui en fait plutôt mention à la section sur la spécificité à laquelle elle est étroitement liée. On doit donc se tourner vers d'autres sources pour avoir une définition formelle de cette qualité: "L'exactitude de l'analyse correspond à la fidélité avec laquelle est traduit le contenu d'un document" (Bertrand-Gastaldy, 1986, p.10). Ainsi, l'indexeur doit s'efforcer de respecter la pensée de l'auteur, c'est-à-dire qu'il ne doit pas introduire de fausses relations entre ces termes d'indexation ou retenir des notions qui ne sont pas véritablement présentes dans le document. La connaissance du domaine est donc une condition essentielle à l'application de l'exactitude.

En plus d'une connaissance insuffisante du sujet, les causes du manque d'exactitude sont associées à un manque d'objectivité au moment du choix des concepts ainsi qu'au moment de la traduction en langage documentaire, à la difficulté de lecture du document, ainsi qu'à la transcription au moment de l'enregistrement (e.g. fautes d'orthographe) (Bertrand-Gastaldy, 1986).

L'application systématique de ce critère par un indexeur devrait probablement être repérable par une plus forte tendance à suggérer des nouveaux termes d'indexation

(candidats) lorsqu'une notion est jugée primordiale et qu'elle ne peut être représentée sous sa forme exacte ou traduite de manière adéquate avec les termes du langage.

Notons également que la confiance des usagers dans le système documentaire repose principalement sur l'exactitude de l'indexation et sur l'impartialité de l'indexeur lors de son jugement de l'exactitude d'un terme.

### Pertinence

Dans son expression la plus simple, la pertinence est caractérisée par le rapport entre un *besoin* et un *document* évalué par une *personne* (Wang et Soergel, 1998). Citons premièrement la définition qu'en donne Bonneville (1989) dans le manuel d'analyse et d'indexation du Réseau québécois d'information sur la communication (RQIC) :

Le descripteur pertinent est le descripteur le plus apte à rendre compte d'un concept contenu dans le document indexé. La pertinence de l'indexation est une qualité essentielle. Si l'indexation est déficiente à cet égard, des concepts (A) sont représentés par des descripteurs inappropriés, c'est-à-dire qui devraient être représentés par d'autres concepts (B). Il s'ensuit que le chercheur désirant retrouver des documents traitant des concepts B repère des documents portant sur des concepts A. Il s'agit du phénomène de bruit documentaire. (Bonneville, 1989, in Patch-Neveu, 1989, p.25)

Rappelons que les taux de rappel et de précision permettent d'évaluer le degré de pertinence au repérage.

La pertinence est un concept central en SI qui pose problème au niveau de sa définition. La littérature dans ce domaine fournit différentes acceptions de la pertinence, aucune ne faisant l'unanimité. Doit-elle être définie en fonction du document, du système, de l'utilisateur, de l'indexeur? En fait, il existe plusieurs types de pertinences et la distinction ténue, en anglais, entre les notions de *pertinence* et *relevance* accentue l'ambiguïté pour une définition précise et unique.

Mizzaro (1997) passe en revue l'ensemble des recherches traitant de ce concept et décrit les différentes approches et l'évolution du concept depuis les quarante dernières

années (principalement de 1958 à 1996). Il est hors de notre propos de faire ici l'historique de cette notion, toutefois il est utile d'en exposer les caractéristiques dominantes pour clarifier la façon (ou les façons) dont l'indexeur peut en tenir compte dans son analyse.

Deux traditions majeures distinguent les travaux sur la pertinence, soit la pertinence liée au propos du document (*topicality, system relevance, relevance at the source*) et la pertinence liée à l'utilisateur (*utility, pertinence, satisfaction*) (Janes, 1991).

La première perspective est orientée vers l'appariement des termes de la question avec les termes d'indexation, c'est-à-dire que le document est jugé pertinent si le mot de la question correspond à un mot présent dans le document; qui a été retenu comme terme d'indexation. Cette perspective étant étroitement liée au critère d'exactitude. Les normes et politiques s'en tiennent généralement à cette première acception, comme le démontre les conseils d'ERIC : "Are the major terms reflected (at least conceptually or by equivalent terms) in the Abstract or Title, so that there may be no question in the user's mind about their relevance?" (ERIC, 1980, p.463). Pour la seconde perspective, qui tient compte de dimensions plus pragmatiques, le document est jugé pertinent si le contenu correspond aux intérêts de recherche de l'utilisateur, c'est-à-dire qu'il remplit le besoin d'information, indépendamment de la présence des mots de la question dans le document (Wang et Soergel, 1998). D'une part, on attribue une certaine objectivité à la pertinence, critère mesurable, inhérent au document et à la question, et d'autre part, on considère le caractère plus subjectif, dynamique et multidimensionnel de la pertinence (Mizzaro, 1997).

À ce propos, plusieurs auteurs (Goffman et Newill, 1966; Foskett, 1970 in Mizzaro, 1997; Wellisch, 1994) soulignent l'importance de distinguer les concepts de *relevance* et de *pertinence*. Un document sera *relevant* pour un utilisateur parce qu'il répond à une question précise, lui fournit de l'information plus détaillée sur un sujet déjà connu, lui permet de clarifier une idée ou d'associer des notions de manière différente. Par contre, la *pertinence* dépend de facteurs subjectifs se distinguant du propos (*aboutness*) du document tels la connaissance préalable du document de l'utilisateur ("j'ai déjà vu ce document"), le niveau de traitement ("c'est trop élémentaire"), l'année de parution ("c'est trop vieux") ou la langue ("je ne lis pas le russe") (Wellisch, 1994).

Toutefois, cette distinction de terminologie ne fait pas l'unanimité. Certains auteurs parlent de *relevance* face au propos du document et de *relevance* pour l'utilisateur (Vickery, 1959; Maron et Kuhns, 1960 in Mizzaro, 1997). D'autres introduisent la notion d'utilité (*usefulness*) reliée aux caractéristiques individuelles de la personne qui évalue l'information (Rees et Schultz, 1966; Regazzi, 1988 in Mizzaro, 1997) rejoignant ainsi la signification de la *pertinence* mentionnée par Wellisch (1994). Par exemple, Soergel (1994) et Maglaughlin et Sonnenwald (2002) soutiennent qu'un document est dit *utile* dans la mesure où l'information fournie n'est pas connue de l'utilisateur et contribue à agrandir ses connaissances sur ce sujet.

Quelle que soit la terminologie employée pour définir ce concept ou la perspective privilégiée pour l'étudier, il est admis qu'en réalité, la pertinence d'un terme ou d'une indexation se situe en bonne partie du côté de l'utilisateur (Green, 1995; Janes, 1991; Lancaster, 1998, Zhang, 2002), "a great deal of information about relevance lies in user-defined criteria (Maglaughlin et Sonnenwald, 2002, p.329). Comme le soutient d'ailleurs Soergel (1994), "different search requests have different requirements, and the quality of the search can be judged only in light of these requirements" (p.590).

On trouve également le point de vue opposé soutenant que la pertinence pouvant être prise en compte et évaluée par l'indexeur au moment de l'analyse se rapporte et se limite au propos du document:

Thus, an indexer can only hope to assign to a document index terms that will indicate the relevance of a topic dealt with in that document. What future users will make of it – its pertinence – will forever be beyond the indexer's ken. (Wellisch, 1994, p.132)

En conséquence, devrait-on se référer à la notion d'*indexabilité* pour rendre compte de manière plus juste de la prise en compte de la pertinence par l'indexeur? La notion d'*indexabilité*, telle que décrite par les politiques d'ERIC, semble correspondre à une évaluation de la *relevance* avant le repérage. Ainsi, un indexeur soucieux de ne retenir que l'information jugée *indexable* devrait tenir compte des recommandations suivantes (ERIC, 1980, p.430-431) :

- Indexer le contenu réel du document et non les implications potentielles.
- Être conscient que l'indexation porte sur des concepts et non sur des mots. Ainsi, ne pas entreprendre une indexation basée exclusivement sur le vocabulaire exact du document, mais bien relever les concepts sous-jacents; un auteur peut utiliser un vocabulaire exagérément exotique, spécialisé, familier, pouvant nuire au repérage.
- La seule présence d'un concept dans le document n'en fait pas nécessairement une information *indexable*. Le concept sera *indexable* s'il s'agit entre autres, de la raison pour laquelle le document est écrit ou si sans ce concept, le document n'existerait pas ou serait altéré de manière significative.
- L'*indexabilité* ou l'importance d'un concept peut être évaluée en fonction de l'ampleur du traitement. La décision repose alors sur le jugement de l'indexeur qui doit développer une intuition (*a feel*) de ce qui peut être considéré comme un traitement suffisant faisant d'un concept une information *indexable*.

En bref, "the indexer should assume the posture of a skeptic who has to be convinced by good evidence of the appropriateness and applicability of any given concept" (ERIC, 1980, p.431).

### Usager

Le fait de ne pouvoir tenir compte de la pertinence à l'étape de l'analyse du propos du document, renvoie à une autre réalité du travail de l'indexeur, soit la difficulté de prendre en compte les besoins potentiels des usagers<sup>3</sup>.

Les études sur l'usager constituent un champ de recherche à part entière en SI. Notre sujet de recherche étant le travail de l'indexeur, nous ne décrivons pas les recherches portant

---

<sup>3</sup> Dorénavant, nous utiliserons l'expression usager(s) pour désigner la dimension comprenant la prise en compte des besoins des usagers, la prise en compte d'un type particulier d'usagers et la référence à l'usager en tant qu'utilisateur de la base de données.

sur les éléments en aval de la chaîne documentaire. Ainsi, l'utilisateur n'est évoqué que dans la perspective de sa prise en compte par l'indexeur.

En plus des normes et politiques qui insistent sur ce point, la majorité des auteurs en SI s'entendent pour dire qu'il est primordial de tenir compte des besoins de la clientèle dans le processus d'indexation mais que ceux-ci, paradoxalement, sont difficilement prévisibles (Bates, 1998; Farrow, 1994; Milstead, 1994; Wellisch, 1994): "...indexing decisions are based on predicted utility, in that an indexer's selection of concepts must bear in mind the questions, as far as these can be known, that may be put to the indexing system" (Farrow, 1994, p.165).

En effet, les usagers d'un centre documentaire forment rarement une clientèle homogène, la nature de leurs besoins variant en fonction de leur domaine d'intérêt, de leur niveau de connaissance, de leur motif d'utilisation du système, etc. Même un centre documentaire qui possède des politiques d'indexation explicites et élaborées concernant les différents types de clientèles et leurs intérêts respectifs ne pourra prévoir l'ensemble des requêtes<sup>4</sup>. Milstead (1994) rappelle qu'il n'existe pas d'utilisateur "générique" ou "absolu" auquel tous les indexeurs se référeraient pour les guider dans leur sélection de termes.

Il semble que toute tentative de la part de l'indexeur pour retenir des termes en prévision des choix de termes des usagers demeure incertaine. Bates (1998) affirme qu'en réalité, cette cohérence indexeur-utilisateur est un leurre puisque leur connaissance du domaine, du langage documentaire ainsi que des règles le gouvernant et du fonctionnement du système ne sont pas et ne seront jamais du même niveau. En effet, leur *expérience* respective avec un système documentaire donné se rapporte à une fonction différente. L'utilisateur doit formuler une requête visant à repérer des références pertinentes, dont il ne connaît pas le contenu *a priori*, tandis que l'indexeur transpose en une liste de termes représentatifs, le contenu d'un document qu'il a sous les yeux. Même avec l'assistance d'un intermédiaire, l'utilisateur éprouve de la difficulté à exprimer clairement et précisément son

---

<sup>4</sup> Notons que la majorité des centres documentaires québécois n'ont pas de politiques écrites concernant les usagers (Bertrand-Gastaldy, 1997 : entrevue personnelle).

besoin d'information (Belkin, 1984; Morris, 1994; Bates, 1998) dû à ses connaissances parcellaires de l'ensemble des composantes du système.

### Approches orientée-document et orientée-usager

Les travaux en SI présentent également le processus global d'indexation selon deux approches: l'indexation orientée-document et l'indexation orientée-usager (Fidel, 1994; Marchionini, 1995; Soergel, 1994). Il est à noter, qu'en théorie comme en pratique, une indexation doit souscrire aux deux approches bien qu'elles soient décrites de manière distincte.

La première approche, centrée sur le document, considère l'indexation surtout comme la représentation du contenu du document, l'indexat servant de substitut au texte original permettant de situer ce document en rapport à un ensemble de documents similaires. Pour cette approche, l'objectif de l'indexation est de représenter (décrire, exprimer, indiquer) le contenu d'un document. Ainsi, le processus d'indexation est considéré selon ces deux composantes principales:

- 1) L'analyse du contenu résultant en la sélection de termes représentatifs du contenu du document;
- 2) La traduction ou l'expression des concepts choisis en termes d'indexation.

Cette approche n'exclut pas l'utilisateur, mais n'en fait pas un élément central du processus. L'utilisateur pourra être décrit en termes généraux, mais la prise en compte d'un utilisateur individuel à la recherche d'information (*information seeker*) n'est pas abordée (Fidel, 1994). D'ailleurs, les normes internationale (ISO, 1985) et nationales (AFNOR, 1978; BSI, 1984) adoptent ce même niveau de description de l'utilisateur, laissant le soin aux centres de documentation d'établir des paramètres plus précis. En résumé, l'approche orientée-document vise en tout premier lieu à exprimer le propos (*aboutness*) d'un document.

La seconde approche insiste sur la fonction de l'indexation qui consiste à faciliter le repérage de documents. Ainsi, l'indexation orientée-usager fait de l'utilisateur son élément pivot en cherchant à prévoir les requêtes. L'indexeur devra choisir ses termes d'indexation

en fonction des termes d'indexation les plus fréquemment utilisés par les usagers pour ce genre de document (Fidel, 1994; Soergel, 1994).

Les tenants de l'approche orientée-document valorisent la saisie "plein texte" de l'information conservant ainsi le vocabulaire de l'auteur afin de ne pas perdre ou occulter la précision des termes originaux lors d'une traduction en vocabulaire contrôlé, bien que ce dernier facilite la manipulation de données. Des recherches plus récentes se rapportent au développement d'outils informatiques visant à concilier les deux approches (usager et document) afin de conserver la richesse du texte original et de rendre le codage, la manipulation et le repérage plus aisés (Bates, 2002; Lancaster, 1998; Shobowale, 1998).

Tout comme la description des critères de qualité dans les normes, l'approche orientée-document n'est décrite qu'en termes généraux. Que suppose cette approche au niveau de l'appréhension du contenu du document? Qu'est-ce que l'indexeur doit traiter pour rendre compte du propos du document, et de quelle manière doit-il en traiter? Ces questions renvoient au débat précédemment décrit concernant la nature ambiguë du *propos* d'un document.

L'approche orientée-usager peut être associée à la règle de sélectivité (Chaumier, 1988) pour laquelle l'indexeur doit évaluer son indexation en fonction de sa pertinence pour l'utilisateur : "si un utilisateur pose une question sur ce concept et que ce document lui soit proposé, serait-il pertinent aux yeux de cet utilisateur?" (Chaumier, 1988, p.56).

La norme ISO (1985) souligne d'ailleurs: "La qualité de l'indexation sera d'autant mieux atteinte que les indexeurs auront un contact direct avec les utilisateurs" (ISO, 1985, p.4), ce qui, dans la pratique, est rarement le cas. On peut alors se demander quelle est la connaissance réelle des indexeurs concernant les besoins d'information de la clientèle et comment peuvent-ils en tenir compte? Dans un pareil cas, ils devront se fier à la description donnée par les politiques du centre ou adopter une approche plus orientée-document lorsque le centre ne possède aucune consigne leur permettant d'indexer en fonction des besoins de la clientèle.

Suite à cette brève revue d'un ensemble d'éléments de connaissance impliqués dans la résolution de l'indexation, il semble clair que les normes et politiques soulignent le caractère essentiel de chaque élément pour la qualité globale de l'indexation sans indiquer, toutefois, la manière dont peuvent être appliqués simultanément ces différents éléments relevant de principes parfois contradictoires. En effet, il ne semble pas y avoir de directives indiquant ce qui doit primer lors de l'analyse.

Notons qu'une description précise de l'application de tous ces éléments s'avère peut-être impossible puisque chaque indexation répond à un contexte précis (politiques d'indexation propres à un centre, nature des documents, langage documentaire, besoins de la clientèle, etc.). Mais quel que soit le contexte, l'indexeur doit faire un usage abondant de son jugement personnel afin de produire une liste de termes la plus "parfaite" possible, c'est-à-dire qui réponde de manière satisfaisante<sup>5</sup> aux exigences de la tâche.

Il est donc justifiable de vouloir examiner l'appréhension de ces éléments de connaissance par différents indexeurs puisque :

- 1) Les normes (et les politiques dans bien des cas) sont suffisamment générales pour laisser une grande place au jugement de l'indexeur.
- 2) Les critères de qualité n'ont pas été examinés du point de vue des processus intellectuels de l'indexeur, c'est-à-dire de la manière dont l'indexeur les définit et les applique.

Puisque l'on suppose que le manque de cohérence interindexeurs peut être causé par une appréhension et une application différentes des critères de qualité et de certains éléments de connaissance relatifs à la tâche (besoins des usagers, pertinence), on doit pouvoir repérer et caractériser l'utilisation de ces divers éléments qui permettent à un indexeur de déterminer que son indexation est satisfaisante. Ainsi, pour vérifier notre hypothèse générale, nous avons choisi d'étudier le travail intellectuel réel des indexeurs

---

<sup>5</sup> Rappelons que le terme satisfaisante renvoie à la notion de *satisficing* développée par Simon et présentée précédemment.

experts en procédant à une expérimentation qui permettra de recueillir des données directement en cours de tâche.

Le chapitre suivant présente en détail les justifications des choix méthodologiques à la base de notre expérimentation.

## CHAPITRE II : MÉTHODOLOGIE

Pour entamer ce deuxième chapitre, nous présentons d'abord la stratégie générale à la base des choix méthodologiques qui ont conduit au plan expérimental. Nous poursuivons avec une description plus détaillée de la méthode choisie pour recueillir les données nécessaires à la vérification de notre hypothèse. Ensuite, sont présentées la situation expérimentale et les justifications liées aux choix des indexeurs, des documents à indexer et des langages documentaires utilisés. Vient ensuite la description de la tâche demandée et du type de données recueillies. Nous terminons ce chapitre par quelques données relatives à la transcription et la segmentation des protocoles.

Dans un premier temps, rappelons notre hypothèse de base: Avec l'expérience, l'indexeur expert a assimilé les normes et règles de l'indexation de manière personnelle, et en conséquence a développé sa propre méthode pour produire une indexation satisfaisante. Par extension, ceci suppose également que l'utilisation des éléments de connaissance ne sera pas identique d'un indexeur à l'autre et que l'identification de ces différences devrait fournir une nouvelle explication à la problématique de la cohérence interindexeurs.

L'indexation étant avant tout une tâche intellectuelle, la seule observation des gestes des indexeurs (parcours du document, consultation du langage documentaire, écriture des termes d'indexation, etc.) ne permet pas d'identifier les stratégies personnelles qui guident les indexeurs dans leur prise de décision. Il est donc nécessaire d'avoir accès aux processus cognitifs des indexeurs pour pouvoir étudier la nature de la variabilité des choix de termes.

Ainsi pour pouvoir vérifier notre hypothèse, il faut élaborer un plan expérimental qui permette de faire ressortir l'utilisation des divers éléments de connaissance en cours de tâche et qui respecte le milieu naturel de la tâche pour que les données recueillies témoignent des processus impliqués dans la tâche réelle.

Pour ce faire, nous employons la méthode classique en sciences cognitives, l'analyse de protocoles verbaux (Newell et Simon, 1972; Ericsson et Simon, 1980, 1984) et

combinons trois types de verbalisations pour nous permettre de recueillir un éventail le plus complet possible de données relatives aux processus cognitifs des indexeurs, soit, la verbalisation concomitante, la verbalisation consécutive sur trace et une évaluation comparée des indexats. Chaque type de verbalisation sert un objectif particulier. Précisons que nous enregistrons les indexeurs sur bande vidéo et audio pour conserver une trace de ces données. Nous décrivons brièvement chaque type de verbalisations ici puisque nous y revenons de manière plus détaillée plus loin dans l'exposé:

- 1) La verbalisation concomitante, également connue sous le nom de "think aloud protocol" ou pensée à voix haute, est recueillie en cours de tâche et permet d'accéder aux processus cognitifs de l'indexeur au moment même où il exécute ses actions (cognitives et physiques). Il s'agit d'une sorte de recueil "en direct" du cheminement cognitif de l'indexeur lié au processus d'indexation. Ce sont les données "brutes" à partir desquelles la verbalisation consécutive sur trace est recueillie. Ajoutons qu'il s'agit d'une méthode reconnue pour ne pas être trop intrusive.
- 2) La verbalisation consécutive sur trace se rapporte aux explications et justifications fournies par l'indexeur sur son propre comportement, c'est-à-dire fournies à partir de l'observation de l'enregistrement vidéo de son indexation. Ce second type de données vient compléter et consolider les premières données en fournissant des explications à un niveau plus abstrait et global. En effet, ces données permettent de recueillir les explications à la base des actions et des décisions de l'indexeur qui ne sont pas nécessairement accessibles ou visibles à partir des seuls protocoles concomitants. En conséquence, ces données devraient permettre également de caractériser le *satisficing* de l'indexeur en relation à son utilisation particulière des divers éléments de connaissance.
- 3) Comme troisième type de données, nous demandons aux indexeurs une évaluation comparée des indexats. En bref, il s'agit de recueillir les commentaires des indexeurs sur leurs propres listes de descripteurs et sur celles produites par les autres indexeurs. Cette évaluation comparée est utilisée en tant que données supplémentaires visant à pousser plus loin la réflexion de l'indexeur en lui demandant d'explicitier ou d'inférer les raisons/stratégies qui ont motivé les indexeurs à faire tel ou tel choix de descripteur.

Quels sont les éléments et conditions qui doivent être réunis dans notre expérimentation pour mettre à jour les stratégies d'indexation des indexeurs et par conséquent, faire ressortir les éléments de connaissance composant le *satisficing* de chaque indexeur?

Premièrement, pour s'assurer que les différences repérées sont tributaires d'une appréhension et d'une application différentes des divers éléments de connaissance cités précédemment, l'expérimentation doit être élaborée de manière à minimiser l'impact des facteurs identifiés comme les plus influents sur la cohérence interindexeurs, soit la connaissance de l'indexation, la connaissance du langage documentaire utilisé et la connaissance du domaine des documents à indexer (Bertrand, 1993; Leonard, 1977; Markey, 1984). En ce sens, l'on doit sélectionner des indexeurs qui ont un profil des plus similaires, c'est-à-dire qui ont acquis une expertise en indexation, qui sont familiers avec un même langage documentaire et qui connaissent le domaine des documents à indexer. Pour remplir ces conditions, il semble pertinent de choisir des indexeurs experts qui travaillent pour un même centre de documentation et de leur faire indexer les mêmes documents, ce qui permettra une comparaison interindexeurs.

Précisons que la connaissance du domaine de l'environnement provient de l'expérience en tant qu'indexeurs dans ce domaine et qu'aucun des indexeurs de notre expérimentation ne possède une formation préalable en environnement. Ce qui est le cas de la plupart des indexeurs qui œuvrent pour un centre de documentation spécialisé. Ils se familiarisent avec le domaine par la lecture des documents qu'ils indexent.

Ainsi, nous avons fait indexer quatre documents par quatre indexeurs experts provenant du même centre de documentation et donc étant également familier avec le langage documentaire<sup>1</sup>.

---

<sup>1</sup> Nous verrons plus loin que deux des indexeurs ne travaillent plus au centre depuis deux ans et justifierons ce choix malgré cette légère différence de profil.

Nous analysons ensuite les processus de résolution des indexeurs à travers l'indexation de plusieurs documents afin de s'assurer que l'identification du *satisficing* d'un indexeur est constant d'un document à l'autre. Chaque indexeur a donc procédé à l'indexation de quatre documents, choisis en fonction du type de documents habituellement indexés par ces indexeurs.

L'approche privilégiée consistant en un examen en profondeur du travail intellectuel de chaque sujet, nous avons privilégié un nombre peu élevé de sujets plutôt que d'avoir à se restreindre à une analyse superficielle des protocoles d'un plus grand nombre de sujets. D'ailleurs, si notre hypothèse s'avère juste, les différences d'utilisation des éléments de connaissance devraient être révélées à partir de l'analyse comparée du travail de quelques indexeurs. Par ailleurs, précisons que le nombre d'indexeurs pour un même centre de documentation se limite à quelques indexeurs; notre choix étant dicté également par cette réalité. Dans le cas qui nous occupe, le bassin d'indexeurs experts disponibles possédant le profil recherché ne regroupe que cinq indexeurs dont quatre ayant accepté de participer à notre expérimentation. De plus, la méthode d'analyse de protocoles verbaux étant longue et ardue, mais qualitativement riche, la quantité de matériel à analyser issu de trois types de verbalisations de l'indexation de quatre documents par quatre indexeurs constitue un corpus largement suffisant pour remplir les objectifs de notre recherche.

En ce qui concerne le langage documentaire, les indexeurs ont utilisé leur thésaurus habituel pour l'indexation de deux documents. Selon notre stratégie de recherche, nous avons jugé utile d'introduire un deuxième thésaurus avec lequel les indexeurs ne sont pas familiers pour mieux démarquer ce qui est lié spécifiquement au thésaurus en termes de stratégies ou d'éléments de connaissance. Plusieurs raisons motivent ce choix.

Ainsi, il nous a semblé pertinent d'utiliser un second thésaurus pour que tous les indexeurs de notre expérimentation soient au même niveau quant à leur connaissance d'au moins un langage documentaire. En effet, nous ne pouvons conclure à une connaissance équivalente du thésaurus Envirodoq chez tous nos participants étant donné que deux d'entre eux n'ont pas utilisé ce thésaurus depuis un certain temps. Il est donc possible que les deux

indexeurs actuels du MENVIQ possèdent des réflexes que les deux autres indexeurs n'ont plus puisqu'il ne s'agit pas du langage que ces derniers utilisent quotidiennement au travail.

L'introduction de ce nouvel élément peut également constituer un "révélateur" de stratégies, en ce sens que confrontés à un vocabulaire nouveau – un thésaurus similaire dans sa structure, mais différent par son vocabulaire – il est possible que les indexeurs aient recours au langage de manière moins automatique, c'est-à-dire que l'application de certaines stratégies et d'éléments de connaissance devienne plus transparente.

Nous sommes conscients que les données ne peuvent plus, en ce cas, être considérées équivalentes à une situation "naturelle". Cependant, le choix de cette variation repose sur la supposition que les indexeurs experts en sont arrivés à une certaine stabilité dans leur utilisation des éléments de connaissance et qu'en conséquence, l'introduction d'un nouveau langage ne devrait pas altérer leur stratégie d'analyse du document, c'est-à-dire leur application des éléments de connaissance par rapport à ce à quoi doit correspondre une indexation satisfaisante. Si notre hypothèse de départ s'avère juste, la représentation du satisficing de ces indexeurs experts devrait demeurer relativement constante. De plus, cette variation nous permettra peut-être de faire ressortir l'influence des contraintes liées au langage documentaire, ce qui pourra ajouter à la profondeur de notre analyse de la tâche.

En somme, pour notre expérimentation, nous tentons de réunir les éléments et conditions qui, nous l'espérons, permettront de fournir une nouvelle explication à la problématique de la cohérence interindexeurs.

Maintenant que le plan méthodologique a été présenté globalement, la suite du chapitre sera consacrée à une description et une justification détaillées de divers éléments de ce plan.

## **1. Analyse de protocoles verbaux**

Depuis une vingtaine d'années, l'analyse de protocoles verbaux constitue une méthode de recherche privilégiée en sciences cognitives qui permet d'accéder aux processus cognitifs

d'un individu. Tout d'abord utilisée pour les études en résolution de problème (Newell et Simon, 1972; Ericsson et Simon, 1984) et par la suite pour l'analyse du travail (Bainbridge, 1979; Hoc et Leplat, 1983; Hoc, 1984), cette approche suppose que dans ce type d'activités, l'individu ne fonctionne pas uniquement par automatismes, mais doit consciemment construire une représentation de la situation et élaborer des stratégies. Ainsi, il est possible d'accéder aux processus cognitifs puisqu'ils sont en partie conscients et gérés par l'individu. En conséquence, il est possible pour l'individu de rendre à voix haute le déroulement de sa pensée.

Puisque nous avons choisi d'utiliser trois types de verbalisations pour notre expérimentation, chacune d'elles sera maintenant présentée. Soulignons qu'une fois recueillies, les verbalisations sont retranscrites sur papier et doivent être codées, c'est-à-dire segmentées en énoncés significatifs. C'est à partir de l'analyse de ces énoncés que les stratégies et procédures des différents indexeurs seront mises à jour et comparées.

### 1.1 Verbalisation concomitante

Les verbalisations concomitantes visent à recueillir des données sur les éléments utilisés par l'individu au moment même de la réalisation de la tâche: connaissances, opérations, stratégies. Pour ce faire, le sujet doit verbaliser sa pensée pendant l'exécution de la tâche. Cette procédure est généralement filmée et permet ainsi de combiner l'analyse des actions à celle des verbalisations. Il est important de souligner que l'objectif de ce type de verbalisations ne vise pas à recueillir l'interprétation que le sujet fait de ses propres processus, mais à accéder au contenu de sa mémoire de travail sans interférer avec la tâche (Ericsson et Simon, 1980, 1984; Newell et Simon, 1972). Ainsi, il est essentiel que l'expérimentateur mentionne et, s'il le faut, rappelle au sujet qu'il n'a pas à justifier ou expliquer ses actions puisque la validité des verbalisations concomitantes s'appuie justement sur la fidélité de leur correspondance avec le processus réel. Une fois les consignes données, l'expérimentateur n'interviendra dans le déroulement de la tâche que pour relancer la verbalisation du sujet qui garderait silence pour une trop longue période.

La procédure de pensée à voix haute doit se faire de manière la plus transparente possible, c'est-à-dire que le sujet ne doit pas mobiliser ses ressources mentales dans le but de produire un discours identique à celui d'une conversation normale. En d'autres termes, le sujet ne doit pas chercher à se faire comprendre. Il est primordial que le sujet parle sans perdre le fil des processus de résolution de sa tâche. En conséquence, le discours est produit avec moins d'attention et prend la forme de phrases souvent incomplètes, d'ellipses et d'expressions non syntaxiques (Breuleux, 1990). C'est le compromis nécessaire pour s'assurer que la trace recueillie témoigne de l'activité intellectuelle réelle du sujet.

Concernant les verbalisations concomitantes, certaines objections ont été soulevées et l'on trouvera dans Ericsson et Simon (1980, 1984), Caverni (1988) et Branch (2000) une discussion sur leurs limites et leurs avantages. Sans présenter le débat, mentionnons toutefois l'objection principale qui suppose que la verbalisation concomitante constitue une deuxième tâche qui interfère dans le déroulement normal de la tâche principale et modifie sa réalisation. Ericsson et Simon (1984) réfutent cette objection en expliquant que la verbalisation concomitante n'interfère pas avec la tâche si le traitement des informations s'effectue sous forme verbale. Ainsi, une tâche ne nécessitant pas un recodage de l'information afin d'être rapportée sous forme verbale ne pose pas le conflit de la double tâche. Les auteurs précisent que l'argument de la double tâche n'est valable qu'à la condition que les tâches nécessitent un mode d'encodage et de décodage autre que verbal, tels la construction d'une représentation imagée ou le traitement routinier des données qui ne laissent qu'une mince trace en mémoire, voire aucune, ainsi que la plupart des activités motrices.

Mentionnons toutefois que certains auteurs (Hoc 1984; Caverni, 1988) distinguent deux types de consignes utilisées pour les verbalisations concomitantes. La consigne libre, du type "pensez à voix haute" qui vise à recueillir le processus de manière fidèle à la condition réelle d'exécution, ainsi qu'à éviter les oublis et les reconstructions. Cette consigne pose le risque que le sujet passe sous silence des éléments utilisés de manière évidente dans sa tâche (par exemple, la poursuite d'un sous-but). À l'inverse, une consigne contraignante visant à identifier des éléments précis (cherchez-vous à remplir tel sous-but?), peut inciter le sujet à affirmer avoir fait ce qu'il n'a pas fait; les verbalisations étant alors

externes au processus en cours d'exécution. Ainsi, la précision de la consigne n'apparaît pas comme un gage de la validité des données recueillies.

D'ailleurs, plusieurs études ont démontré que la verbalisation concomitante modifie la performance selon que la consigne requiert du sujet la justification de chaque action (Gagné et Smith, 1962 in Caverni, 1988) ou la verbalisation de ses hypothèses (Bower et King, 1967, in Caverni, 1988), tandis que d'autres études (Karpf 1973; Brehmer 1974, in Caverni, 1988) ne trouvent aucune différence de performance.

Afin de recueillir le processus de pensée le plus identique à la tâche réelle, nous optons pour une consigne libre du type "pensez à voix haute" décrite par Ericsson et Simon (1980, 1984), qui ne requiert aucune justification ou explication de la part du sujet pendant l'exécution de la tâche.

Quelques études (Bertrand, 1993; Endres-Niggemeyer, 1990; Tomonori, 1983) ont eu recours à l'analyse de verbalisations concomitantes pour décrire le processus d'indexation. Nous avons vu au chapitre précédent que l'utilisation de ce type de verbalisation a permis à Bertrand (1993) d'identifier la stratégie d'indexation privilégiée par les indexeurs experts, sans toutefois permettre d'expliquer le faible taux de cohérence de ce groupe, c'est-à-dire de mettre à jour les raisons sous-jacentes aux choix de termes différents.

Bien que la verbalisation concomitante semble être la source d'information la plus directe concernant les processus cognitifs d'un individu, elle demeure cependant incomplète, en ce sens qu'elle ne permet pas d'appréhender toutes les composantes d'un processus. En effet, les protocoles verbaux ne peuvent être utilisés pour délimiter les connaissances d'un sujet à propos de la tâche exécutée. Si le sujet ne mentionne pas un élément dans le protocole, cela ne constitue pas la preuve qu'il ne connaît pas cet élément. En fait, il est admis que la plupart des gens peuvent penser plus rapidement qu'ils ne parlent, ainsi, seulement une partie de toute leur activité cognitive est rapportée (Bainbridge, 1979; Caverni, 1988). Pour cette raison, nous avons cru nécessaire de compléter les données des protocoles concomitants par une exploration plus directe de la connaissance du sujet.

Puisque nous visons à expliquer les différences entre indexeurs experts, nous devons avoir accès à un ensemble de données plus subtiles qui permettent de rendre compte des éléments qui guident les décisions des indexeurs. À cette fin, nous proposons de faire converger deux autres instruments fondés eux aussi sur les verbalisations du sujet, soit la verbalisation consécutive sur trace et une évaluation comparée des listes produites.

## 1.2 Verbalisation consécutive sur trace

Comme il a été mentionné précédemment, les verbalisations concomitantes imposent des limites à l'analyse. En effet, les données rendues par la verbalisation concomitante peuvent passer sous silence l'utilisation de certains éléments, particulièrement lorsque le recours à ces éléments provient d'une expertise ou de contraintes non visibles au moment de la résolution.

Afin d'éviter cette limitation, certains chercheurs (Caverni, 1988; Hoc et Leplat, 1983) ont recours à la verbalisation consécutive sur trace<sup>2</sup> qui consiste à visionner avec le sujet la trace de son comportement (en l'occurrence, l'enregistrement vidéo) et à l'interroger sur son interprétation des processus en cours. Ceci donne alors accès à tout un autre ensemble de données qui permettent à l'analyste de corroborer ses inférences faites à partir du protocole concomitant (quel est le but poursuivi au moment X? pourquoi tel terme est-il rejeté?), en plus de rendre compte de certaines stratégies globales ou paramètres pris en compte par le sujet qui ne sont pas forcément verbalisés. La validité de ces données peut être partiellement évaluée en comparaison des données issues des protocoles concomitants.

Ce deuxième type de verbalisations pose toutefois des risques pour la validité des données recueillies. L'argument fréquemment évoqué pour réfuter leur validité repose sur la possibilité que le sujet fasse état de connaissances qu'il n'a pas utilisées lors de l'exécution de la tâche, résultant en une reconstruction de la procédure. Ceci peut être

---

<sup>2</sup> Dans la littérature, l'expression verbalisation rétrospective est également utilisée pour désigner ce type de verbalisation.

associé aux difficultés pour le sujet de se rappeler ce qui s'est réellement passé. Ces difficultés seraient plus fortement ressenties lorsque l'exécution de la tâche demande beaucoup de temps et qu'un long intervalle sépare l'exécution de la tâche et le recueil de la verbalisation consécutive (Hoc et Leplat, 1983). Un second argument concerne la consigne donnée ou le genre de questions posées par l'expérimentateur. Il se peut que les questions de l'expérimentateur ne correspondent pas à la représentation de la tâche que possède le sujet, le conduisant alors à produire des réponses hors du propos, c'est-à-dire qui ne sont pas en rapport avec l'exécution même de la tâche, ou à formuler des réponses en fonction des attentes supposées de l'expérimentateur (Bainbridge, 1979).

Selon le modèle proposé par Ericsson et Simon (1984), il est possible d'éviter ces risques en utilisant une procédure qui tienne compte du fonctionnement de la mémoire. Idéalement, ce type de verbalisations doit être produit dès que la tâche est complétée. L'exécution de la tâche laissant une trace durable en mémoire, le sujet peut accéder à cette trace en partie à travers sa mémoire à court terme (MCT), ou la récupérer en mémoire à long terme (MLT), pour ensuite la verbaliser. Leur modèle stipule également que le sujet a facilement accès à la trace cognitive de l'activité et qu'il peut en témoigner sans que l'expérimentateur lui indique la nature précise des informations devant être rapportées. Dans ces conditions, la consigne de verbalisation peut être générale : "de rapporter tout ce dont vous vous rappelez avoir pensé durant l'exécution de la tâche".

Par ailleurs, Ericsson et Simon (1984) suggèrent d'opter pour une consigne générale pour le recueil des verbalisations consécutives plutôt que de restreindre le sujet à un ensemble prédéfini d'alternatives. En effet, une question formulée en termes très spécifiques risque d'influencer la réponse du sujet, à tel point qu'il devient impossible d'affirmer que sa réponse témoigne du processus de pensée en cours de résolution. Par exemple, prenons les trois questions/réponses suivantes:

1) Q. Avez-vous utilisé le sous-but X?

R. Oui.

2) Q. Avez-vous utilisé un sous-but? Si oui, lequel?

R. Oui, j'ai utilisé le sous-but X.

3) Q. (consigne générale): à quoi pensiez-vous à ce moment?

R. Je voulais parvenir à X,... lorsque j'ai atteint X...

Les deux premières questions peuvent en effet remettre en question la validité des réponses en tant que représentation du traitement cognitif réel, tandis que la troisième est une question ouverte qui ne restreint pas le sujet et par conséquent sous-tend une plus grande validité de la réponse.

D'autre part, Ericsson et Simon (1984) insistent sur la nécessité de s'assurer que le contenu des verbalisations rapportées a été utilisé au moment de la résolution de la tâche et qu'il ne provient pas d'un traitement subséquent ou d'une connaissance antérieure utilisée dans un contexte similaire. Il est possible d'éliminer cet artefact en demandant au sujet de ne rapporter que les détails réellement pris en compte concernant l'épisode ou l'action spécifique sur lequel ou laquelle porte le questionnement. Ceci permet d'éviter que le sujet remplace les détails dont il ne se souvient pas par une information dont il dit "s'être sûrement servi".

Concernant notre expérimentation, il est nécessaire d'identifier les éléments de connaissances (critères de qualités, prise en compte des besoins des usagers, etc.) réellement utilisés par les indexeurs pour les choix de termes et les stratégies privilégiées pour accomplir leur tâche. Ainsi, nous avons demandé aux sujets de fournir des explications pour un épisode, une action ou une décision particulière, la consigne prenant alors la forme d'un questionnement ponctuel du genre, "qu'avez-vous pensé à ce moment?", "pouvez-vous m'expliquer ce qui se passe?", ou "pourquoi avez-vous fait cette action ou pris cette décision?".

Un dernier facteur rarement évoqué doit également être pris en considération pour l'évaluation de la validité des verbalisations, soit l'honnêteté du sujet face au témoignage de ses propres processus intellectuels. L'expérimentateur doit clairement indiquer au sujet que l'expérimentation à laquelle il participe ne consiste pas à évaluer sa compétence, sa

productivité ou à cerner les erreurs, mais bien à permettre au chercheur d'avoir accès à la dimension inaccessible du travail de l'indexeur, que lui seul peut révéler. Ainsi, il est primordial que ce qui est rapporté par l'indexeur corresponde exactement aux traitements réels effectués en cours de tâche.

Quant à la quantité d'informations générée par ce type de verbalisations, Ericsson et Simon (1984) précisent qu'aucun chercheur ne s'est plaint d'une surabondance de données. Au contraire, il semble que les sujets verbalisent généralement moins d'informations que ce qui est souhaité par les expérimentateurs.

Finalement, bien que les études de Hoc et Leplat (1983) démontrent que les verbalisations consécutives sur trace présentent les meilleures garanties de validité, les auteurs suggèrent toutefois de ne pas utiliser la verbalisation comme seul indicateur du fonctionnement cognitif, mais de chercher à confronter les données fournies avec d'autres données ne provenant pas de protocoles verbaux. La verbalisation consécutive pourra alors être complétée par un questionnaire ou une entrevue servant à appuyer l'identification de la représentation de la tâche et de l'utilisation des divers éléments de connaissance chez les sujets.

### 1.3 Évaluation comparée des indexats

Nous avons recueilli aussi un troisième type de données qui consiste pour les indexeurs en une évaluation comparée des listes de descripteurs produites par les autres indexeurs. Les indexeurs ont commenté leurs propres listes ainsi que celles produites par les autres indexeurs.

Cette évaluation de l'ensemble des listes produites nous est apparue comme une source de données complémentaires potentiellement riche pour consolider la nature des utilisations des différents éléments de connaissance identifiées à partir de l'analyse des verbalisations concomitantes et consécutives. Ainsi, ces données supplémentaires servent deux objectifs. D'une part, elles servent à vérifier si les indexeurs se réfèrent aux mêmes

éléments de connaissance évoqués en verbalisations concomitantes et consécutives pour justifier leurs choix de descripteurs. D'autre part, nous voulons vérifier si les indexeurs peuvent inférer les stratégies particulières des autres indexeurs à partir de l'examen des listes de descripteurs. Qu'ils puissent ou non reconnaître les stratégies réellement utilisées par les autres indexeurs, ils devraient néanmoins être en mesure de fournir une explication pour les différences interindexeurs. L'ensemble de ces données est donc utilisé pour étayer la caractérisation du *satisficing* de chaque indexeur puisque nous supposons que les explications fournies devraient se rapporter en partie aux différents éléments de connaissance.

De plus, à la suite de l'examen des listes des autres indexeurs, chaque indexeur peut apporter des modifications à ses listes et fournir les raisons sous-jacentes à ces changements. Nous voulons voir l'influence de la consultation des autres indexations sur leur propre travail. En ce sens, nous pouvons vérifier si les indexeurs tendent vers un plus grand accord une fois qu'ils ont pris connaissance du travail d'autres indexeurs et quels arguments sont évoqués pour justifier les modifications.

Ce recueil s'est effectué environ deux mois après l'indexation. Plusieurs raisons motivent ce choix méthodologique.

Premièrement, il nous semble pertinent de procéder à ce recueil après une certaine période afin que les indexeurs ne puissent se rappeler leurs verbalisations concomitantes et les réponses fournies en verbalisation consécutives sur trace. Ainsi, tout recoupement de la mention d'éléments de connaissance avec les premières verbalisations pour la justification du rejet ou du choix de descripteurs viendra vraisemblablement confirmer la composition du *satisficing* d'un indexeur.

De plus, deux mois nous apparaît un délai suffisant pour que les indexeurs aient pris un recul par rapport à leurs indexations et pour qu'ils puissent être tout autant critique par rapport à leur propre travail qu'à celui de leurs pairs. Il est évidemment souhaitable qu'ils puissent effectuer ces comparaisons en toute objectivité, sans chercher à argumenter plus en faveur de leurs indexations.

Précisons que nous avons interrogé les indexeurs uniquement sur les listes produites pour les deux documents indexés avec Envirodoq, soit *Étang aéré* et *Biogaz*. Nous avons limité cette étape à deux documents, soit une demie journée, puisque l'expérimentation de base, deux jours complets, est relativement exigeante pour les indexeurs. Par ailleurs, nous croyons que les données issues de cette évaluation seront suffisantes pour reconnaître l'utilisation des éléments de connaissance propre à chaque indexeur et pour permettre aux indexeurs d'identifier ou inférer un processus particulier chez leurs pairs.

En conclusion, la combinaison de ces trois méthodes devrait nous permettre d'obtenir un ensemble de données qui va de données brutes et ponctuelles à des interprétations de niveau plus général afin de saisir à la fois l'application particulière à chaque indexeur de l'ensemble des éléments de connaissance relatifs à la tâche d'indexation et leur représentation globale de l'indexation. À notre connaissance, la tâche d'indexation n'a jamais encore été examinée à l'aide de ces trois méthodes simultanées.

#### 1.4 Transcription, segmentation et codage des protocoles

L'un des principaux objectifs de l'analyse de protocoles verbaux consiste à identifier les prises de décisions des sujets, les éléments qu'ils utilisent pour prendre leurs décisions et l'organisation de ses éléments (procédures, stratégies) dans la résolution globale de la tâche (Bainbridge, 1979).

Les verbalisations recueillies ne peuvent être exploitées directement sous leur forme initiale (enregistrements audio et vidéo) et doivent être transcrites intégralement sur papier avant d'être analysées. Une fois les verbalisations transcrites, la première étape d'une analyse de protocole consiste à segmenter le discours du sujet en énoncés significatifs. Un énoncé correspond le plus souvent à une phrase, une proposition, une idée se rapportant à un même élément (action, objet, sous-but, état, etc.). Certains indices de changement de traitement d'un élément sont marqués dans le discours par une variation de l'intonation de la voix, par des hésitations, des pauses, etc. (Caverni, 1988). Relativement simple à

effectuer, la procédure de segmentation permet généralement un niveau élevé de fiabilité (Ericsson et Simon, 1984). L'étape suivante consiste à coder ces énoncés, c'est-à-dire à les répertorier parmi l'ensemble des catégories prédéterminées qui sont généralement affinées en cours d'analyse des données.

Évidemment, la dégradation du discours des pensées à voix haute entraîne des difficultés pour l'analyse des protocoles concomitants, et plus particulièrement pour le codage. En effet, l'information fournie par ces protocoles ne permet pas d'identifier avec certitude le recours à un élément de connaissance particulier pour justifier une décision. Il s'ensuit que l'assignation d'une catégorie à un énoncé doit être tributaire du contenu explicite des protocoles et des inférences faites par l'analyste à partir de sa propre connaissance de la tâche. Mais une telle analyse demeurerait incomplète pour l'identification des traitements cognitifs puisque certaines inférences effectuées par l'analyste peuvent s'avérer inexactes et mener à de fausses interprétations.

Ainsi, le recours à la verbalisation consécutive sur trace permet de pallier aux informations non explicitement formulées, puisque le sujet complète les premiers protocoles en fournissant explications et justifications pour ses actions, décisions, etc.

Cependant, les formes verbales employées par les sujets pour exprimer les divers éléments de connaissance qu'ils utilisent peuvent ne pas correspondre à la forme exacte employée dans la norme et peuvent varier d'un sujet à l'autre. Donc, une partie de l'analyse des protocoles doit reposer sur l'examen en profondeur des informations verbalisées pour identifier l'élément de connaissance réellement utilisé au-delà de la structure de surface des énoncés produits par le sujet.

### 1.5 Élaboration des catégories

Le but du codage consiste à répertorier les énoncés des protocoles sous forme d'opérations, de traitements effectués sur des éléments pour pouvoir décrire et caractériser le comportement cognitif d'un ou de plusieurs individus. Il s'agit donc d'élaborer une

codification qui permette de rendre compte de tous les énoncés, sinon du plus grand ensemble possible.

L'élaboration des catégories servant au codage et à la segmentation pourra se faire en deux temps. Premièrement, certaines catégories pourront être déterminées à partir d'une analyse préalable de la tâche ainsi qu'en fonction de l'objectif de la recherche. Ces catégories sont de nature générale et regroupent habituellement les éléments qui sont invariablement utilisés pour la résolution de la tâche. Pour les problèmes bien définis par exemple, il s'agit de catégories issues de l'énoncé même du problème qui peuvent regrouper des objets, opérateurs, buts, etc. (Bainbridge, 1979).

Dans un deuxième temps, le développement des catégories réside dans l'examen d'un premier protocole afin d'extraire un ensemble de catégories qui puisse rendre compte de manière exhaustive des énoncés. Il s'agit ensuite de coder un second protocole en appliquant les catégories élaborées à partir du premier et de raffiner la codification en précisant certaines catégories ou en ajoutant des catégories qui ne se trouvaient pas dans le premier protocole. L'ensemble des catégories ne peut être développé que par itération. Il faut donc coder quelques protocoles de plus avec la même codification jusqu'à ce que le système de notation se stabilise.

On doit également porter une attention particulière au choix d'un niveau d'analyse approprié. Bien qu'il soit souhaitable que la codification puisse rendre compte du plus grand nombre d'éléments utilisés par les sujets, il ne faut cependant pas tomber dans l'excès d'exhaustivité en ajoutant une nouvelle catégorie pour tout nouvel élément. D'ailleurs, un grain d'analyse extrêmement fin risque de rendre impraticable toute généralisation, ne faisant ressortir que des idiosyncrasies et masquant ainsi les informations pertinentes nécessaires à l'atteinte des objectifs de la recherche. Ericsson et Simon (1984) classent les énoncés qui ne témoignent d'aucune des catégories identifiées en tant que "commentaires". Il est important de conserver ces énoncés non répertoriés par le système de notation car ils peuvent se révéler indicateurs d'une particularité du comportement d'un sujet et servir alors à le caractériser par rapport aux autres sujets.

Rappelons que la rigidité ou la souplesse du système de notation dépend du type de tâche examinée et surtout de l'objectif visé par la recherche. L'analyse d'un problème bien défini visant à mettre au point une codification qui permette d'automatiser la procédure de codage (Breuleux, 1990; Newell et Simon, 1972) sera nécessairement plus rigide que l'analyse d'une tâche moins bien définie visant à mettre à jour les processus intellectuels utilisés pour l'accomplissement de la tâche (Farrow, 1995; Tomonori, 1983).

Nous présenterons maintenant la situation expérimentale de notre recherche afin de spécifier l'utilisation particulière que nous faisons des divers éléments méthodologiques.

## **2. Situation expérimentale**

Afin de vérifier notre hypothèse, nous avons élaboré une situation expérimentale qui permet de recueillir les informations essentielles à l'analyse de l'indexation et de cerner les différences interindexeurs.

Pour s'assurer que les différences repérées sont bien tributaires d'une appréhension et application différentes des éléments de connaissance, l'on doit être en mesure de cerner de manière précise leur rôle et leur influence dans le processus d'indexation des indexeurs. En conséquence, rappelons que les sujets et le contexte de l'expérimentation (type de document, langage documentaire, niveau d'indexation, etc.) ont été déterminés de manière à minimiser l'impact des facteurs ayant été identifiés comme les plus influents sur la cohérence interindexeurs, soit la connaissance de l'indexation, la connaissance du langage documentaire utilisé, et la connaissance du domaine des documents à indexer (Bertrand-Gastaldy, 1986; Leonard, 1977; Markey, 1984).

## 2.1 Choix des sujets

Pour pouvoir démontrer que les variations interindexeurs sont dues à une appréhension et une application différentes des éléments de connaissance reliés à la tâche d'indexation, on doit sélectionner des indexeurs possédant des profils des plus similaires possible. Il est donc préférable de choisir des indexeurs experts travaillant pour un même centre de documentation. En effet, de tels indexeurs ont acquis une expérience de l'indexation, utilisent le même langage documentaire, sont familiers avec le(s) domaine(s) et les types de documents à indexer, en plus de connaître les politiques d'indexation du centre. Ces indexeurs experts ont vraisemblablement atteint une certaine stabilité dans leur procédure d'indexation. En conséquence, les différences interexperts de choix de termes sont susceptibles d'être attribuables à l'utilisation d'une stratégie d'indexation personnelle. Soulignons que l'on entend par expertise en indexation, d'une part, la connaissance de règles formelles à suivre qui sont enseignées (quelles parties du document consulter, comment tenir compte de la clientèle, quel niveau d'analyse privilégier, etc.) ainsi que les conseils de mise en forme. D'autre part, il s'agit également de l'expérience en indexation qui comprend le développement de l'expertise par la pratique (la maîtrise des politiques du centre, la connaissance de la clientèle, la connaissance des divers types de documents de la base de données, etc.).

Rappelons que les taux de cohérence les plus élevés relevés par les études antérieures (Leonard, 1977; Bertrand, 1993) sont observés chez des experts utilisant un même thésaurus. Ainsi, nous centrons nos choix sur des indexeurs experts travaillant avec un thésaurus.

Par ailleurs, le choix de nos sujets relève d'une décision d'ordre pragmatique autant que théorique. En effet, la présente étude s'inscrit dans un projet de recherche plus vaste effectué conjointement avec une équipe de chercheurs en sciences de l'information dont les recherches portent sur le travail des indexeurs et la base de données bibliographiques Envirodoq du Centre de documentation et de renseignements du Ministère de

l'environnement du Québec (MENVIQ)<sup>3</sup>. Nous nous sommes assurés la collaboration des indexeurs experts de ce centre qui possèdent le profil requis pour notre expérimentation (connaissance de l'indexation, du domaine, du thésaurus et du fonctionnement interne du centre). Cependant, seulement deux indexeurs œuvraient au sein du MENVIQ<sup>4</sup> au moment de notre cueillette de données. La population d'indexeurs correspondant au profil choisi étant très restreinte, nous avons sollicité la participation d'indexeurs experts ayant travaillé récemment au MENVIQ. Deux des trois indexeurs ayant quitté le centre depuis environ deux ans ont accepté de participer à notre recherche. Ceci porte à quatre le nombre de sujets pour notre expérimentation.

### *2.1.1 Profil des indexeurs*

Les deux indexeurs du centre (E2 et E3) travaillent au MENVIQ depuis deux et trois années respectivement, tandis que les deux autres (E1 et E4) y ont travaillé pour une durée d'un an et deux ans<sup>5</sup>. Tous ces indexeurs ont acquis une formation en indexation (technique documentaire ou bibliothéconomie), sans qu'aucun d'entre eux en ait fait son champ de spécialisation. Cela signifie qu'ils connaissent tous les principes de base de l'indexation contenus dans la norme internationale (ISO, 1985).

Quant à leur expérience générale de l'indexation, ils peuvent tous être considérés experts, E1 possédant quinze ans d'expérience, E2 trois ans, E3 quatre ans et E4 vingt-trois ans.

Bien que ces indexeurs soient tous experts, rendant possible la comparaison de leurs résultats, leur expertise n'est cependant pas identique. De fait, les deux indexeurs qui ne travaillent plus au MENVIQ depuis deux ans devront utiliser des connaissances dont ils ne se sont pas servis depuis un certain temps. Ce facteur devra être pris en considération lors de l'analyse des résultats puisque certains choix de termes ou certaines décisions pourront

---

<sup>3</sup> La refonte ministérielle de 1995 a réuni deux ministères, le MENVIQ s'appelant dorénavant le Ministère de l'Environnement et de la Faune du Québec. L'expérimentation a eu lieu avant cette fusion, ainsi nous conserverons le titre du MENVIQ pour désigner ce ministère dans la présente recherche.

<sup>4</sup> Il est à noter qu'en raison de restriction budgétaire, le nombre d'indexeurs est passé de quatre à deux.

<sup>5</sup> Précisons ici que le chiffre attribué à chaque indexeur correspond à l'ordre de passation de l'expérimentation.

être attribuables à une connaissance inégale ou insuffisante du contexte. Mais puisque ce que l'on cherche à mettre à jour est l'appréhension et l'utilisation différentes d'éléments de connaissance impliqués dans la tâche d'indexation, par des indexeurs experts, il est possible que l'inégalité apparente de leur expérience récente avec le thésaurus et la base de données Envirodoq n'ait aucune incidence sur l'application effective de leur stratégie personnelle. Rappelons que nous supposons que chaque indexeur a atteint une certaine stabilité dans sa stratégie globale d'indexation et qu'en conséquence, la représentation et l'application de son *satisficing* devraient être relativement constantes.

## 2.2 Choix des documents

Le choix de documents doit répondre principalement à deux contraintes: 1) le domaine et le type de documents doivent être familiers aux indexeurs et; 2) les documents ne doivent pas avoir déjà été traités par ces indexeurs. Il s'agit de documents dans le domaine de l'environnement. Plusieurs types de documents sont traités par les indexeurs du MENVIQ: congrès, exposés de congrès, études contractuelles, mémoires, rapports, etc. Nous avons opté pour des documents dont la structure et la nature du contenu demandent un traitement suffisamment élaboré de la part des indexeurs qui permettra de faire ressortir l'application de leurs stratégies d'indexation. En effet, l'indexation superficielle (descripteurs principaux) de documents de vulgarisation par exemple, ne saurait fournir le type de données dont nous avons besoin. En l'occurrence, nous favorisons les rapports, les études contractuelles ou les mémoires. Notons que la politique officieuse du MENVIQ concernant les documents volumineux est d'attribuer une moyenne de dix à douze descripteurs principaux (DE1) et secondaires (DE2), c'est-à-dire une vue d'ensemble par les descripteurs principaux et une représentation plus détaillée pour les aspects particuliers par les descripteurs secondaires.

Pour s'assurer que les documents utilisés pour notre expérimentation n'ont pas été indexés préalablement par les indexeurs, nous avons emprunté des documents internes produits par des firmes spécialisées dans la recherche et développement en environnement.

Quatre documents ont été retenus, dont deux rapports de recherche et deux études de faisabilité, ce qui correspond aux types de documents régulièrement indexés au MENVIQ<sup>6</sup>.

L'indexation de quatre documents devrait fournir un éventail suffisant d'éléments de connaissance utilisés et permettre de caractériser la procédure d'indexation des indexeurs. Également, nous limitons à quatre le nombre de documents indexés par chaque indexeur en fonction des exigences de la méthode d'analyse de protocoles mentionnées précédemment.

### 2.3 Choix des langages documentaires

Les indexeurs experts utiliseront deux thésaurus lors de l'expérimentation. Le thésaurus Envirodoq, élaboré et utilisé par le centre de documentation et de renseignements du MENVIQ; et Infoterra, un thésaurus également relatif à l'environnement produit par Infoterra, le système international de référence aux sources de renseignements sur l'environnement du Programme des Nations Unies pour l'environnement.

Nous avons déjà explicité les raisons qui nous motivent à avoir recours à un second thésaurus. Rappelons qu'il s'agit d'un choix méthodologique visant à mieux cerner les stratégies et les éléments de connaissance liés au thésaurus, en plus de faire ressortir les contraintes qu'exerce le langage documentaire sur l'application des divers éléments de connaissance.

En outre, l'utilisation du thésaurus Infoterra ne devrait pas poser de problème majeur pour les indexeurs puisque la présentation de ses informations est comparable à celle d'Envirodoq en plusieurs points. D'ailleurs, le choix d'Infoterra comme deuxième thésaurus s'est imposé de lui-même. Parmi les rares thésaurus se rapportant exclusivement à l'environnement, Infoterra est celui qui partage le plus de caractéristiques avec Envirodoq. Le Tableau VI présente de manière synthétique la structure du contenu de ces thésaurus<sup>7</sup>.

---

<sup>6</sup> Nous avons pris connaissance des types de documents les plus couramment indexés par les indexeurs du MENVIQ lors d'un entretien téléphonique avec le responsable du centre de documentation.

<sup>7</sup> Rappelons que la section 1.2 du premier chapitre présente les types de relations utilisées, par ces deux thésaurus, pour la présentation du vocabulaire.

Tableau VI. Caractéristiques des thésaurus Envirodoq et Infoterra.

	ENVIRODOQ	INFOTERRA
<b>TOTAL DES TERMES</b>	<b>1596</b>	<b>2099</b>
• Descripteurs (termes préférentiels)	1486	1561
• Non-descripteurs (termes non préférentiels)	110	261
• Descripteurs géographiques	-----	200
• Descripteurs fonctionnels	-----	77
<b>RELATIONS</b>		
• Équivalence	Employé (EM) Employé pour (EP)	Employé (EM) Employé pour (EP)
• Hiérarchique	Terme générique (TG) Terme spécifique (TS)	Terme générique (TG) Terme spécifique (TS)
• Associative	Terme associé (TA)	Terme associé (TA)
• Définition	Note d'application (NA)	Note d'application (NA)
<b>PRÉSENTATIONS</b>		
• Alphabétique	Tous les termes en ordre alphabétique	Tous les termes en ordre alphabétique
• Sommaire des catégories	-----	Les 21 catégories et sous-catégories d'après leur numéro de code. Aucun terme particulier.
• Catégories et sous-catégories	-----	Tous les termes en ordre alphabétique par catégorie et sous- catégorie.
• Numéro de code	-----	Tous les termes d'après leur numéro de code.

Les thésaurus sont comparables en termes du nombre de mots répertoriés pour décrire le domaine (1596 pour Envirodoq et 2099 pour Infoterra), des types de relations (TG, TS, TA, EM, EP, NA) et de la présentation des termes dans une liste alphabétique ou lexicale. En fait, il s'agit de thésaurus dont la structure de base est la même, soit une liste de mots accompagnée des relations unissant les termes.

On remarque une différence pour les descripteurs géographiques (DEgéo) et fonctionnels. Précisons que le thésaurus Envirodoq ne contient pas de descripteurs géographiques. Toutefois, les indexeurs inscriront en tant que descripteurs géographiques les termes importants d'un document qui se rapportent aux aspects géographiques<sup>8</sup>. Quant aux descripteurs fonctionnels, Infoterra les regroupe dans une section permettant ainsi d'identifier la langue de travail, la nature de la source (gouvernementale, commerciale, etc.) et les conditions d'accès (frais négociables, sur la base d'échanges, etc.).

L'emploi du thésaurus Infoterra dans 140 pays justifie la nécessité de cette liste particulière de soixante-sept descripteurs pour préciser la langue de travail, la nature de la source, etc. Les descripteurs fonctionnels d'Envirodoq font partie de la liste de termes sans distinction particulière. Les documents intégrés à la base de données Envirodoq sont de langue française ou anglaise, et un endroit sur le bordereau sert à indiquer les documents de langue anglaise.

La différence principale des deux thésaurus se retrouve dans la présentation de leur contenu. Envirodoq est constitué d'une liste alphabétique, tandis qu'en plus d'une liste alphabétique, Infoterra comprend trois autres représentations (sommaire des catégories, catégories et sous-catégories, et numéro de code). Il est possible que les indexeurs experts habitués à indexer uniquement à partir d'une liste alphabétique éprouvent quelques difficultés à se servir des autres représentations. Nous devons donc porter une attention particulière à l'étape de consultation du thésaurus et noter les différences de stratégies en comparaison avec l'utilisation du thésaurus habituel (Envirodoq).

## 2.4 Déroulement de l'expérience

### 2.4.1 Description de la tâche demandée

Tous les indexeurs indexent les quatre documents l'un à la suite de l'autre, dans le même ordre. Nous croyons qu'un effet d'ordre est peu probable avec des indexeurs

---

<sup>8</sup> Le MENVIQ possède d'ailleurs un ouvrage de référence précisant la forme exacte à employer pour la plupart

professionnels. Les deux premiers documents sont analysés avec le thésaurus connu, Envirodoq, et les deux autres avec Infoterra. Notre hypothèse de base suppose une forte stabilité dans la procédure et l'utilisation des éléments de connaissance par les indexeurs experts. De même, il nous apparaît peu probable que l'ordre d'utilisation des langages documentaires influence le processus au point de recueillir un tout autre ensemble de données si Infoterra avait été utilisé en premier.

Ce sont les versions imprimées des thésaurus qui sont utilisées pour l'expérimentation. Notons que malgré l'existence d'une version ordinolinguage d'Envirodoq, les indexeurs du MENVIQ favorisent la version papier dû à la lenteur du système utilisé pour faire fonctionner le logiciel.

#### *2.4.2 Production de listes*

Tous les indexeurs doivent produire deux listes de termes. Une première liste, que l'on nomme feuille brouillon, feuille de travail ou liste brouillon, sert d'aide-mémoire tout autant que de liste de termes potentiellement intéressants pour l'indexation finale. Soulignons que nous n'avons imposé aucune contrainte quant à l'utilisation de cette feuille. Les indexeurs ont pour consigne d'utiliser cette feuille comme bon leur semble, c'est-à-dire comme ils le font en situation normale. En effet, dans la pratique de l'indexation, la majorité des indexeurs se servent de feuilles brouillon pour prendre des notes autant à la première étape de prise de connaissance du contenu que lors de l'élaboration des choix de descripteurs.

La production de cette première liste permet également de conserver une trace de la représentation initiale du document et de l'isoler de la liste finale de descripteurs afin de faire ressortir les contraintes liées à l'utilisation d'un vocabulaire contrôlé, en l'occurrence les deux thésaurus. Dans une certaine mesure, cette liste pourra également servir d'indicateur de la connaissance du thésaurus pour chaque expert lorsqu'ils utiliseront le thésaurus Envirodoq.

La deuxième liste est constituée des descripteurs retenus pour l'indexation, c'est-à-dire les résultats de l'analyse. C'est à partir de ces listes que les calculs de cohérence sont effectués. La feuille de descripteurs que nous leur fournissons comporte une section pour les descripteurs principaux et une autre pour les secondaires. Les indexeurs pouvaient ajouter des divisions telles que descripteurs géographiques et candidats descripteurs s'ils les jugeaient appropriées pour la représentation des documents. En fait, ils devaient procéder comme lorsqu'ils indexent pour le MENVIQ et inscrire les termes retenus en fonction du champ approprié (descripteurs principaux, secondaires, géographiques et candidats descripteurs). Soulignons que dans le contexte habituel, les indexeurs procèdent de cette manière, c'est-à-dire qu'ils ont en main le document à indexer, le thésaurus et des feuilles brouillon.

Cependant, pour conserver une trace du résultat de l'indexation, nous demandons aux indexeurs d'inscrire les descripteurs sur une feuille plutôt que sur la notice informatisée. Notons que seulement les indexeurs du MENVIQ auraient pu avoir accès à la base de données durant l'expérimentation puisqu'elle s'est déroulée sur les lieux de travail.

### *2.4.3 Réalisation de la tâche*

Avant de procéder à l'expérimentation, l'expérimentateur lit les consignes avec le sujet. Tous les indexeurs ont pris connaissance des consignes générales suivantes :

#### **Consignes générales pour le recueil des protocoles concomitants**

L'expérience consiste à observer comment s'élabore la pensée de quelqu'un qui indexe un document. Puisque mon but est de mieux comprendre comment vous faites pour faire l'indexation, j'ai besoin de savoir à quoi vous pensez pendant que vous indexez. Ce qu'il y a de particulier à l'indexation que je vous demande de faire, c'est que vous ferez de la verbalisation concomitante à la tâche, c'est-à-dire qu'il s'agit pour vous d'exprimer à voix haute le contenu de votre pensée.

En somme, il s'agit de dire tout haut ce que vous pensez. Exprimez tout: les suppositions, les pensées qui sont en rapport avec la tâche ou les pensées qui n'ont aucun rapport avec la tâche. Parlez de façon aussi continue que possible.

Il est très important de **penser, de lire et d'écrire** à voix haute. Vous ne devez pas expliquer ou justifier ce que vous faites. Ce que je veux, c'est simplement que vous pensiez à voix haute. Laissez sortir votre petite voix intérieure. Bref, agissez comme si vous étiez seul(e) et que vous vous parliez à vous-même. Essayez d'en arriver à dire vraiment ce que vous pensez sur le moment plutôt que de penser pour un bout de temps et de décrire ensuite ce à quoi vous avez pensé. Si vous demeurez silencieux (silencieuse) trop longtemps, je vais vous demander "À quoi pensez-vous?" ou "Tout haut s'il-vous-plaît", pour vous rappeler de verbaliser.

J'enregistre le tout sur cassettes vidéo et audio et ce sont ces enregistrements que j'utiliserai pour ma recherche. Ils sont réservés à mon usage personnel.

Avez-vous des questions?

Premièrement, je vous demande de faire l'indexation de quatre documents. Il y a une seconde étape pour laquelle nous visionnerons ensemble votre indexation et je vous poserai des questions.

#### *2.4.4 Entraînement à la verbalisation*

L'expérimentation débute par une pratique à la verbalisation. Puisque les gens ne sont pas habitués à faire ce genre d'exercice, l'expérimentation est généralement précédée d'une phase d'entraînement. On demande alors à la personne de pratiquer cette technique sur un ou plusieurs autres problèmes. Puisqu'il s'agit d'un exercice, l'expérimentateur peut interrompre la personne dans sa résolution afin d'ajuster la façon de verbaliser du sujet, de même le sujet peut interrompre sa tâche afin de poser des questions. Dans notre cas, les tâches d'entraînement consistait à résoudre le problème de la tour de Hanoï avec trois et quatre disques, et à effectuer l'indexation en vocabulaire libre d'un article de vulgarisation scientifique d'une page.

### **2.5 Recueil des protocoles concomitants**

Une fois que le sujet s'est familiarisé avec la procédure de verbalisation, l'expérimentateur lui donne le premier thésaurus, la feuille brouillon pour la prise de notes, la feuille devant contenir la liste de descripteurs ainsi que le premier document à indexer.

Précisons que l'expérimentation s'est déroulée sur les lieux de travail pour les indexeurs du MENVIQ (E2 et E3), sur les lieux de travail actuel d'un des deux autres indexeurs (E4) et dans un laboratoire de recherche universitaire pour le quatrième indexeur (E1). Le temps requis pour l'indexation des quatre documents varie entre deux heures et deux heures trente par indexeur.

Aucune limite n'a été imposée pour la durée de l'indexation ou pour le nombre de descripteurs attribués. Les indexeurs devaient procéder comme s'ils indexaient pour le MENVIQ.

Les deux premiers documents indexés avec le thésaurus Envirodoq s'intitulent:

- 1) Étude de sites d'implantation d'étangs aérés (Société Québécoise d'Assainissement des eaux, 1990).
- 2) Élimination du biogaz. Sélection d'une technologie et évaluation des impacts (Tecsult, 1992).

Les deux documents indexés avec le thésaurus Infoterra s'intitulent:

- 3) Compostage des boues résiduelles urbaines (Ministère de l'environnement, 1984).
- 4) Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal (Desbois, 1985).

Afin que l'indexeur prenne connaissance de la structure et du contenu du thésaurus Infoterra, une période variant de cinq à dix minutes d'examen du thésaurus précède l'indexation des documents.

### *2.5.1 Condition particulière*

Soulignons que nous avons altéré légèrement les conditions habituelles d'indexation de nos indexeurs experts en ne leur permettant pas de se référer à la base de données, au guide toponymique ou à certains dictionnaires spécialisés.

En effet, les indexeurs du MENVIQ peuvent consulter à l'occasion les notices dans la base de données qui ont été produites pour des documents similaires à celui devant être indexé pour s'aider dans leur choix de descripteurs. Puisque seuls les deux indexeurs actuels du MENVIQ auraient eu accès à ces notices en consultant la base de données, nous avons préféré soustraire cette option afin que la situation expérimentale soit identique pour tous les indexeurs et surtout que les résultats obtenus demeurent comparables.

Il en va de même pour l'utilisation du guide toponymique servant à repérer la forme exacte à employer pour les descripteurs géographiques. Pour l'expérimentation, les indexeurs inscriront donc le terme géographique tel qu'il apparaît dans le document.

De plus, il arrive parfois que les indexeurs consultent des dictionnaires spécialisés pour connaître la définition exacte d'un terme du document et en évaluer la justesse pour l'indexation. Ces dictionnaires n'étant disponibles qu'au MENVIQ, nous avons jugé pertinent de les mettre de côté pour l'expérimentation.

## 2.6 Recueil des protocoles consécutifs

Le recueil des protocoles consécutifs a suivi immédiatement l'indexation des quatre documents. Pour des raisons de validité des données, les indexeurs n'ont pas été interrogés après chaque indexation puisque la connaissance du type de questions posées lors de la verbalisation consécutive auraient pu influencer leur analyse des documents subséquents.

Pour cette étape, chaque indexeur a visionné l'enregistrement vidéo de son propre comportement afin de fournir explications et justifications à l'expérimentateur. Les protocoles concomitants sont présentés dans le même ordre que les documents ont été indexés. En plus de la trace vidéo, l'indexeur peut consulter les listes de termes et de descripteurs produites pour cette indexation ainsi que le document indexé. Deux types de questionnements ont été utilisés.

Un premier questionnaire relativement ouvert, du type "qu'avez-vous pensé à ce moment?", "pouvez-vous m'expliquer ce qui se passe?", etc., vise à recueillir les explications et justifications concernant une situation ponctuelle (action, décision, épisode, stratégie, etc.) se rapportant uniquement à l'indexation visionnée. Afin de s'assurer de la validité de ces verbalisations, nous portons une attention particulière à la formulation des questions afin qu'elles ne contiennent pas de suggestion de réponse. Aussi, nous demandons régulièrement aux sujets si leurs explications et justifications témoignent de ce qu'ils ont pensé au moment précis de l'action ou de la décision sur laquelle porte le questionnaire.

En parallèle au questionnaire se rapportant aux situations ponctuelles des protocoles concomitants, nous interrogeons également les indexeurs sur leur procédure habituelle de travail, leurs stratégies, les éléments de connaissance privilégiés, etc. Cet ensemble de données permettra de confronter et d'appuyer les informations recueillies à partir de traitements plus ponctuels. Par exemple, lorsque l'indexeur mentionne l'utilisation d'un élément de connaissance pour l'aider dans sa prise de décision, nous lui demandons s'il s'agit d'une stratégie qu'il utilise fréquemment pour régler ce genre de situation. De même, s'il aborde la prise en compte de l'usager, nous le questionnons alors de manière plus directe sur son utilisation précise de cette dimension.

L'intégration d'un deuxième niveau de questionnaire à l'étape des verbalisations consécutives fournit une mise en situation à l'indexeur qui lui permet de préciser la manière dont il utilise les divers éléments de connaissance pour une situation donnée et de préciser les différents contextes dans lesquels cette utilisation peut varier. Ainsi, la combinaison de deux niveaux de questionnaire (sur la trace et sur le processus habituel) permet d'accéder à des explications d'une grande subtilité qui sont primordiales pour caractériser la stratégie d'indexation de chaque indexeur et parvenir à expliquer les différences interindexeurs.

L'entente initiale prise avec les indexeurs consistait en une période de deux jours consécutifs pour la réalisation de l'expérimentation. Ainsi, le recueil des protocoles consécutifs sur trace a nécessité une journée et demie. Une première partie a été recueillie à la suite de l'indexation des quatre documents et ce deuxième ensemble de données a été

complété la journée suivante. Rappelons que, malgré la charge de travail considérable demandée aux indexeurs pour cette expérimentation, il était impératif que le recueil des protocoles consécutifs suive immédiatement la tâche effectuée pour que la trace mentale soit la plus fraîche et accessible possible.

## 2.7 Évaluation comparée des indexats

L'évaluation comparée consistait pour les indexeurs à commenter les listes de descripteurs produites par les autres indexeurs ainsi que leurs propres listes dans le but de faire ressortir les éléments composant leur *satisficing* qui leur permettent de juger de la pertinence d'une indexation. À la suite de l'examen de l'ensemble des listes, l'indexeur pouvait apporter des modifications à ses listes et devait fournir les explications pour les changements effectués.

Nous avons formulé un questionnaire servant de base pour diriger l'entrevue de l'évaluation comparée. Dans une première étape, l'indexeur commente sa propre indexation. On lui demande ensuite de commenter les indexats produits par les autres indexeurs et une dernière étape consiste à recueillir ses commentaires sur l'ensemble des indexations. C'est lors de cette dernière étape que l'indexeur a la possibilité d'apporter des modifications à ses listes de descripteurs. Notons que l'on conserve également les listes produites à la suite des modifications, lesquelles seront utilisées pour calculer la cohérence interindexeurs et vérifier si les indexeurs arrivent à un plus grand accord après l'observation des indexats de leurs pairs.

Le questionnaire de base est reproduit ici.

### Questions pour l'évaluation comparée des indexats.

#### 1) L'indexeur révise sa propre indexation.

a) J'aimerais que tu regardes ton indexation de ce document environ 5 minutes afin de te rafraîchir la mémoire. (L'indexeur peut consulter le thésaurus, le document et sa liste de descripteurs placés devant lui.)

b) Peux-tu me commenter ton indexation?

c) Avec un deuxième regard, est-ce que cette liste te semble ok?

d) Est-ce que tu te rappelles la distinction que tu faisais entre descripteurs principaux et secondaires?

(Cette question sera utilisée pour relancer l'indexeur s'il n'a aucun commentaire)

e) Avais-tu des critères de sélection (pour descripteurs principaux et secondaires)?

## 2) L'indexeur examine et commente les autres listes de descripteurs (l'une après l'autre).

Comme tu le sais, j'ai demandé à d'autres indexeurs d'indexer ce même document. Les choix de descripteurs ne sont pas pareils et j'aimerais que tu commentes ces différences. J'aimerais avoir ton opinion sur chacune des indexations.

### Commentaires sur chaque indexation:

a) Peux-tu commenter cette indexation? (question ouverte, pour savoir spontanément ce qu'il en pense)

b) Peux-tu commenter les termes? Dans un deuxième temps, en comparant avec tes termes? (s'il ne l'a pas fait)

c) À quoi l'indexeur peut avoir pensé pour choisir chacun de ces descripteurs?

d) Pourquoi, à ton avis, cet indexeur a-t-il choisi ces descripteurs? Que voulait-il faire?

e) Peux-tu me commenter la distinction descripteurs principaux et secondaires? (dans le même sens que les questions précédentes; c'est-à-dire, à quoi l'indexeur a pensé pour faire cette distinction)

### Commentaires sur chaque indexation en comparaison avec la sienne:

f) Si l'on compare cette indexation avec la tienne, que peut-on en dire?

g) Comment expliquerais-tu les ressemblances et différences des choix de descripteurs?

## 3) Commentaires généraux de l'indexeur sur l'ensemble des indexations.

a) Qu'est-ce qui explique, selon toi, la différence de ces listes? As-tu des idées, des hypothèses qui pourraient m'éclairer?

b) Crois-tu que les listes de descripteurs différeraient plus ou se ressembleraient plus, s'il n'y avait pas l'utilisation d'un thésaurus?

c) Maintenant que tu as examiné les quatre indexations, comparé ton indexation à chacune d'elles et commenté les différences, as-tu l'impression que si tu avais à réindexer le document tu reprendrais les mêmes descripteurs? Ou crois-tu que tu modifierais ton indexation? Si oui, quelles modifications apporterais-tu et pourquoi?

## 3. Données recueillies

Les données recueillies se divisent en trois groupes: les listes, les protocoles et les données de l'évaluation comparée des indexats.

### 3.1 Listes

Au total, quarante listes ont été produites, soit dix listes par indexeur, réparties de la manière suivante: quatre listes brouillon ou feuilles de travail, quatre listes de descripteurs et, pour l'évaluation comparée, deux listes modifiées de descripteurs.

### 3.2 Protocoles

Trente-deux protocoles ont été recueillis, dont seize protocoles concomitants et seize protocoles consécutifs. Il y a huit protocoles par indexeur, dont quatre protocoles concomitants issus de l'indexation des documents et quatre protocoles consécutifs.

### 3.3 Évaluation comparée des indexats

L'entrevue associée à l'évaluation comparée des indexats a duré entre deux et trois heures par indexeur. Quatre entrevues ont été recueillies.

Globalement, l'expérimentation a duré deux jours et demi pour chaque indexeur, pour un total entre treize et seize heures par indexeur. Au total, cela constitue une soixantaine d'heures d'enregistrement à transcrire et à analyser. Le Tableau VII énumère les divers protocoles et entrevues produits par un indexeur pour l'expérimentation.

Tableau VII. Liste des protocoles et entrevue recueillis pour un indexeur associés à la durée de chaque étape.

Étape	Nombre et type de protocoles	Durée
Indexation des quatre documents	4 protocoles concomitants	Entre 2 et 2.5 heures
Explications et justifications des actions pour l'indexation des quatre documents	4 protocoles consécutifs sur trace	Entre 9 et 10 heures
Évaluation des indexats produits par tous les indexeurs pour les documents <i>Étang aéré</i> et <i>Biogaz</i>	1 Entrevue: Évaluation comparée des indexats	Entre 2 et 3 heures

### 3.4 Matériel audiovisuel

Toutes les données ont été enregistrées sur bande audio et vidéo simultanément, sauf pour le recueil de l'évaluation comparée qui n'a été enregistré que sur bande audio<sup>9</sup>.

L'utilisation d'enregistrements vidéo de la résolution de l'indexation permet d'identifier les différentes activités concordantes aux verbalisations, telles que lecture, écriture, et d'avoir une trace des comportements non verbalisés, tels que la consultation des différentes parties des documents, les différentes consultations dans le thésaurus, les annotations autres que l'écriture de termes ou descripteurs (ratures, signes particuliers, etc.). La complémentarité des enregistrements audio et vidéo permet ainsi de repérer le moment précis où toute activité a eu lieu.

### 3.5 Dispositif expérimental

L'aspect technique de l'expérimentation a été élaboré de manière à ce que les sujets puissent rapidement faire abstraction des appareils et ainsi se consacrer à leur tâche comme ils le font en situation réelle. Pour ce faire, la caméra posée sur un trépied est hors du champ de vision de l'indexeur et située en retrait du côté gauche. Ainsi, en utilisant le zoom, il est possible de capter ce qui est écrit, lu et consulté. À ce propos, l'expérimentateur mentionne au sujet que l'on filme principalement en plans rapprochés puisque l'on doit voir le détail des objets manipulés (sections du document, page du thésaurus, termes écrits, etc.), ce qui implique qu'on ne voit que les avant-bras du sujet et presque jamais son visage. Quant à l'enregistrement audio, un micro-cravate est utilisé afin d'entendre clairement toutes les paroles du sujet qui extériorise sa pensée comme s'il se parle à lui-même. Ces enregistrements audio sont utilisés pour la transcription des protocoles concomitants.

---

<sup>9</sup> Nous avons mentionné dès notre première rencontre avec les indexeurs que l'expérimentation nécessite l'utilisation d'une caméra vidéo et d'une enregistreuse puisque cette technique peut paraître intrusive pour

L'expérimentateur est assisté d'un technicien qui voit aux enregistrements audio et vidéo, prend en note la durée de chacune des étapes et manipule la caméra lorsque des gros plans sont nécessaires. Ainsi, l'expérimentateur peut concentrer son attention sur la tâche effectuée par l'indexeur. De plus, cette division des tâches réduit les déplacements physiques au minimum afin de ne pas déconcentrer l'indexeur. Une fois les consignes données, l'expérimentateur s'assoit à l'écart de l'indexeur de manière à pouvoir observer son comportement et prendre des notes.

Pour le recueil des verbalisations consécutives sur trace, un moniteur est placé sur la table devant l'indexeur et l'expérimentateur afin qu'ils visionnent les protocoles concomitants. La caméra est située derrière eux de manière à enregistrer simultanément les commentaires et les explications concernant l'indexation ainsi que le déroulement de l'indexation. Les enregistrements audio sont également utilisés pour la transcription des protocoles consécutifs.

L'étape de l'évaluation comparée des indexats est enregistrée sur bande audio uniquement. Nous avons donc pris soin de bien identifier verbalement chaque liste de descripteurs présentée au moment du recueil de ces données.

#### **4. Transcription des protocoles recueillis**

Les protocoles concomitants issus de notre expérimentation ont été transcrits intégralement. Les verbalisations consécutives sont ensuite insérées à l'intérieur des transcriptions des protocoles concomitants afin de rassembler en un tout continu le processus d'indexation et les explications concernant chacune des étapes, décisions, évaluations, choix de termes, etc. Cette disposition des données simplifie le travail d'analyse puisque chaque épisode, décision, etc. est relié directement à l'explication qui s'y rapporte.

---

certaine personne. De ce fait, les indexeurs ont pu prendre leur décision de participer à l'expérimentation en sachant qu'ils seraient filmés et que leurs paroles seraient enregistrées.

Puisque le texte des protocoles ne permet pas de distinguer les activités d'écriture, de lecture et de traitement recueillies en verbalisations concomitantes, nous utilisons une typographie différente pour chacune d'elles:

- Activité de lecture: texte souligné
- Activité d'écriture: **texte en gras**
- Activité de traitement: texte normal  
(pensée à voix haute)

Pour permettre la distinction entre les verbalisations des protocoles concomitants et les verbalisations consécutives, nous avons encadré les verbalisations concomitantes. Le texte en italique représente des précisions insérées dans l'un ou l'autre des types de verbalisations par l'expérimentateur dans le but de faciliter la compréhension du lecteur. Également, les mots faisant l'objet d'un traitement sont mis en italique. Quant aux textes des évaluations comparées, ils ont été compilés séparément. C'est à partir de ces documents ainsi transcrits que s'effectue l'analyse.

#### 4.1 Segmentation des protocoles et catégories

Une première segmentation s'est imposée d'elle-même par l'insertion des verbalisations consécutives à l'intérieur des protocoles concomitants. Les blocs d'énoncés sont ainsi constitués d'une portion du processus d'indexation suivi des explications de l'indexeur concernant cette portion. Mais une telle segmentation, associée à un arrêt de la bande vidéo, doit nécessairement être raffinée à l'aide de catégories se rapportant aux procédures et connaissances utilisées pour l'indexation.

Deux ensembles de catégories peuvent être établis afin de guider les premières analyses. L'un des ensembles sert à identifier les dimensions plus procédurales de la tâche, tels que les différentes étapes de l'indexation (prise de connaissance du contenu, extraction des informations essentielles, traduction en descripteurs), le recours à la macrostructure du document, la consultation des différentes parties du document, ainsi que la consultation du

thésaurus. L'autre ensemble se rapporte aux différents éléments de connaissance utilisés pour le traitement de l'information, tels que l'évaluation de choix de termes. Comme l'on ne peut prédire l'appréhension et l'utilisation de tous les éléments par les indexeurs, les catégories préétablies pour la codification des protocoles demeurent générales et devront être raffinées par l'analyse de plusieurs protocoles. L'ensemble des catégories suivant vise donc à identifier les éléments de connaissance principaux potentiellement utilisés pour la production d'une indexation satisfaisante, tels que sélectionnés au contexte théorique: 1) critères de qualité soulignés dans les normes auxquels doit répondre une indexation, soit l'exhaustivité, la précision, l'exactitude; 2) prise en compte des besoins des usagers; 3) pertinence (ou *indexabilité*).

L'on cherche également à identifier le rôle de ces éléments dans le processus d'indexation de chaque indexeur en rapport avec des éléments plus contextuels, tels que la prise en compte des politiques d'indexation (explicites ou implicites) du centre de documentation, le fonds documentaire, les contraintes imposées par le langage documentaire, ainsi qu'en fonction d'une approche globale de l'indexation orientée plus document ou plus usager.

Soulignons que l'identification de ces éléments est facilitée par une mention explicite de l'utilisation, tandis qu'une mention implicite devra être identifiée par l'examen du contexte et par la comparaison avec des situations similaires pour lesquelles les mêmes éléments ont été utilisés, mais ne sont pas verbalisés par l'indexeur ou sont exprimés en termes moins exacts.

## **5. Calculs de cohérence**

Les taux de cohérence interindexeurs sont calculés à partir des listes de descripteurs produites pour les quatre documents ainsi que des listes de la réindexation de deux documents en entrevue complémentaire, c'est-à-dire pour l'évaluation comparée des indexats.

Rappelons que la cohérence interindexeurs est une mesure quantitative du degré d'accord entre plusieurs indexeurs en fonction de termes choisis individuellement et séparément par chacun d'eux pour représenter le contenu d'un même document (Rolling, 1981).

En plus des pourcentages établis à partir de l'ensemble des descripteurs de chaque liste, nous calculons les taux de cohérence interindexeurs pour chaque sous-ensemble de descripteurs: Descripteurs principaux (DE1), descripteurs secondaires (DE2), descripteurs géographiques (DEgé), candidats descripteurs (candidats).

En plus de permettre la comparaison de nos résultats avec ceux d'études antérieures, les taux de cohérence interindexeurs obtenus pour les différents sous-ensembles fourniront peut-être quelques pistes à suivre lors de l'examen des protocoles pour expliquer les divergences de choix de descripteurs. Nous calculons également la cohérence de chaque indexeur par rapport au groupe. Notons que la production des listes brouillon (feuilles de travail) n'étant régie par aucune contrainte, nous n'établissons pas de taux de cohérence interindexeurs pour ces listes.

Pour calculer le degré d'accord entre deux indexeurs, il s'agit de diviser le nombre de termes communs affectés par deux indexeurs, par le nombre total de termes affectés par les deux indexeurs (Rolling, 1981). La comparaison doit se faire par paire d'indexeurs. La moyenne des pourcentages calculés par paire permet d'établir le taux de cohérence pour l'ensemble du groupe pour un document donné. Ainsi, la formule permettant d'établir ce rapport est :

$$CO = 2C / (A + B) \quad \text{où} \quad CO = \text{Cohérence interindexeurs}$$

A = nombre total de termes de l'indexeur A.  
B = nombre total de termes de l'indexeur B.  
C = nombre de termes communs aux indexeurs A et B.

Le résultat varie entre 0 et 1, où 1 équivaut à 100 % de recouvrement.

Pour les besoins de la démonstration, nous présentons l'ensemble des données ayant servi au calcul de la cohérence interindexeurs, soit le nombre de termes total par indexeur et le nombre de termes communs à deux indexeurs pour l'ensemble des six paires d'indexeurs, pour l'indexation du premier document de notre expérimentation (*Étude de sites d'implantation d'étangs aérés*).

INDEXEURS	A	B	C	CO
E1 – E2	10	12	6	54 %
E1 – E3	10	5	3	40 %
E1 – E4	10	21	6	38 %
E2 – E3	12	5	5	58 %
E2 – E4	12	21	9	54 %
E3 – E4	5	21	5	38 %

Ainsi, tel que souligné pour la paire E1 et E2, un indexeur A ayant sélectionné 10 termes et un indexeur B ayant sélectionné 12 termes, le nombre de termes communs C entre les deux indexeurs étant de 6 termes, la cohérence interindexeurs est calculée en utilisant l'équation présentée plus haut<sup>10</sup>. Le chiffre obtenu étant multiplié par 100 pour avoir un pourcentage, on obtient :

$$CO = 2C / (A + B) = 2 \times 6 / (10 + 12) = 12 / 22 = 0.54 \times 100 = 54\%$$

Pour obtenir le taux de cohérence (TCO) de l'indexation pour un document indexé par quatre indexeurs, il faut premièrement calculer la cohérence pour les six paires d'indexeurs et ensuite faire la moyenne de ces six paires.

$$TCO = (CO_{E1-E2} + CO_{E1-E3} + CO_{E1-E4} + CO_{E2-E3} + CO_{E2-E4} + CO_{E3-E4}) / 6$$

$$TCO = (54\% + 40\% + 38\% + 58\% + 54\% + 38\%) / 6 = 47\%$$

En ce qui concerne le taux de cohérence de chaque indexeur par rapport au groupe, il s'agit de calculer la moyenne de la cohérence obtenue pour toutes les paires d'indexeurs impliquant un indexeur donné. Ainsi, si l'on veut établir le degré de cohérence de E2 par

---

<sup>10</sup> Les pourcentages obtenus sont arrondis à l'unité.

rapport au groupe pour l'indexation du document *Étude de sites d'implantation d'étangs aérés*, il faut établir le calcul suivant<sup>11</sup>:

$$TCO_g = (CO_{E1-E2} + CO_{E2-E3} + CO_{E2-E4}) / 3$$

$$TCO_g = (54\% + 58\% + 54\%) / 3 = 55\%$$

Précisons que pour le calcul de cohérence, nous considérons comme équivalents les termes dont la forme est au pluriel ou singulier, les acronymes et les abréviations ainsi que leurs formulations complètes. Cette flexibilité quant à la détermination de l'équivalence des termes provient du contexte même de l'expérimentation. Rappelons que pour notre expérimentation, les indexeurs écrivent leurs termes ainsi que leurs descripteurs sur papier, mais qu'en situation réelle, ils devraient transcrire ou écrire directement leurs descripteurs dans la base de données. Nous supposons que s'ils avaient eu à effectuer cette étape, les indexeurs auraient écrit le terme au long et auraient employé le nombre (généralement le singulier) prescrit par la politique du centre.

Par contre, l'utilisation d'un terme non préférentiel dans une liste de descripteurs plutôt que la forme prescrite, est considérée comme une erreur et ce terme n'est pas considéré comme terme équivalent au descripteur utilisé par un autre indexeur. Les omissions de mots dans les listes de descripteurs sont également identifiées en tant qu'erreur et ne peuvent faire partie de la compilation de termes communs.

En somme, nous optons pour un calcul de cohérence basé sur l'appariement de la forme exacte des descripteurs des listes finales qui, en situation réelle, seraient disponibles dans la base de données et disponibles également pour l'utilisateur.

---

<sup>11</sup> TCO<sub>g</sub> = taux de cohérence d'un indexeur par rapport au groupe.

## CHAPITRE III: RÉSULTATS

Le chapitre que nous abordons est consacré à l'analyse des résultats. La première partie se rapporte aux analyses quantitatives et expose ce que les taux de cohérence et la comparaison des listes de descripteurs permettent d'expliquer à propos de la tâche d'indexation et des variations interindexeurs. La deuxième partie, relative aux analyses qualitatives, est centrée sur les verbalisations des indexeurs et vise à fournir une explication cognitive nouvelle au problème de la cohérence interindexeurs. En examinant la procédure d'indexation, l'appréhension et l'utilisation des éléments de connaissance liés à la tâche, nous tenterons d'identifier la stratégie d'indexation personnelle de chaque indexeur, ou le *satisficing*, ce qui devrait permettre en retour de mieux comprendre les variations interindexeurs.

### 1. Analyses quantitatives

Après un bref rappel du profil des quatre indexeurs, un résumé décrit chaque document et certaines particularités d'indexation propres à chacun sont présentées. Les analyses débutent par une comparaison interindexeurs en fonction des listes de descripteurs produites pour les quatre documents et des taux de cohérence. Ces analyses servent de point de départ pour la caractérisation des indexeurs.

Précisons que nous avons jugé pertinent de reporter la présentation des résultats relatifs à l'évaluation comparée des indexats, c'est-à-dire au troisième type de verbalisations, à une section ultérieure (2.6) étant donné qu'il s'agit de données complémentaires. En effet, rappelons que ces données ne serviront pas à la caractérisation du *satisficing*, mais bien à la vérification ou validation de ce qui aura déjà été identifié lors des analyses des deux premiers types de verbalisations.

## 1.1 Identification des indexeurs

Les indexeurs sont identifiés par la lettre E, pour expert, suivi du chiffre correspondant à l'ordre dans lequel les indexeurs ont effectué l'expérimentation.

E1 = a travaillé deux années au MENVIQ et a quitté depuis deux ans.

E2 = travaille depuis deux années au MENVIQ.

E3 = travaille depuis trois années au MENVIQ.

E4 = a travaillé une année au MENVIQ et a quitté depuis deux ans.

## 1.2 Présentation des documents

Débutons par un résumé et une caractérisation de chaque document selon certaines particularités rencontrées par les indexeurs lors de l'indexation afin de situer le lecteur qui n'a pas pris connaissance du contenu des documents<sup>1</sup>.

- Document *Étang aéré*<sup>2</sup>

Le premier document, indexé avec le thésaurus Envirodoq, s'intitule *Étude de sites d'implantation d'étangs aérés* (Société québécoise d'assainissement des eaux [SQAE], 1990) et compte vingt-six pages. Ce document traite de l'évaluation de cinq emplacements potentiels pour la construction d'une usine d'épuration dans le but de traiter les eaux usées de trois municipalités. Les avantages et inconvénients de chaque site sont présentés, et aucun site ne ressort comme le plus approprié.

Concernant certaines particularités de l'indexation de ce document, l'on doit mentionner que les sites sont évalués en fonction de plusieurs types de contraintes dont ne peut rendre compte complètement le vocabulaire du thésaurus. Le titre figurant sur la page couverture est différent de celui de la page titre.

---

<sup>1</sup>Le lecteur pourra prendre connaissance des annexes D, E, F, G, présentant les principales parties des documents consultées par les indexeurs afin d'avoir une meilleure représentation du contenu des documents et par conséquent, du contexte des éléments traités par les indexeurs.

<sup>2</sup>Dorénavant, chaque document est désigné par un mot plutôt que par le titre complet afin d'alléger le texte.

- Document *Biogaz*

Le second document, indexé avec le thésaurus Envirodoq, s'intitule *Élimination du biogaz. Sélection d'une technologie et évaluation des impacts* (Tecscult, 1992). Il contient quatre-vingt pages et quarante-deux pages d'annexes constituées de tableaux. Il s'agit d'un rapport préliminaire concernant une étude d'impact. Ce document présente la démarche suivie pour choisir une technologie adéquate permettant de détruire les biogaz produits par le site d'enfouissement du Centre de Tri et d'Élimination des Déchets de la Carrière Miron. Les impacts de la technologie privilégiée sont évalués en fonction de certaines contraintes environnementales et des préoccupations du public.

L'indexation de ce document peut être caractérisée par le fait que le terme central, *biogaz*, ne se trouve pas dans le thésaurus. La disposition de l'information sur la page titre et la couverture crée une certaine confusion quant à l'identification du titre exact du document.

- Document: *Compostage*

Le troisième document a pour titre *Compostage des boues résiduaires urbaines* (Ministère de l'environnement, 1984). Il est indexé à l'aide du second thésaurus, Infoterra. Il est constitué de cent pages, plus vingt-quatre pages d'annexes et comporte plusieurs tableaux, photos et figures. En plus de décrire en quoi consiste le compostage, ce document présente les avantages environnementaux et économiques du compostage. Il présente une revue détaillée des différents procédés de compostage des boues et décrit les applications de certaines de ces techniques en France.

Plusieurs des termes essentiels retenus par les indexeurs ne se trouvent pas dans le thésaurus sous leur forme exacte. Les descripteurs apparentés sont constitués d'expressions composées qui ne correspondent pas nécessairement au sens du terme initial contenu dans le document.

- Document: *Énergie*

Le quatrième document, également indexé avec le thésaurus Infoterra, s'intitule *Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal* (Desbois, 1985). Il compte cent une page et cinquante-trois pages d'annexes composées de tableaux, figures et schémas. Il s'agit d'un rapport de recherche présenté pour l'obtention d'une maîtrise en science de l'environnement. Présenté comme une étude de faisabilité, ce document fait état des problèmes reliés à l'eau, à la cuisson des aliments et à l'éclairage au village de Ndia Goureye et propose des solutions basées sur l'utilisation d'énergies renouvelables, notamment le soleil, le vent et le biogaz.

Le titre de ce document est très explicite et résume pratiquement à lui seul l'essentiel de son contenu. La majorité des notions représentatives du contenu se trouvent dans le thésaurus.

### 1.3 Taux de cohérence interindexeurs

Le Tableau VIII permet d'observer que les taux de cohérence interindexeurs obtenus pour l'indexation des quatre documents varient entre 28 % et 47 %. Soulignons que les taux de cohérence doivent être pris uniquement à titre indicatif pour illustrer que la cohérence n'est pas très élevée et que les taux obtenus s'apparentent aux taux obtenus de recherches antérieures.

Tableau VIII. Taux de cohérence interindexeurs pour les quatre documents.

Documents	Taux de cohérence interindexeurs
<i>Étang aéré</i>	47 %
<i>Biogaz</i>	28 %
<i>Compostage</i>	40 %
<i>Énergie</i>	44 %

Le taux de cohérence le plus élevé obtenu par nos indexeurs experts est de 47 %, ce qui s'apparente aux résultats obtenus par Bertrand (1993) pour ses groupes d'indexeurs experts (42 % et 45 %).

Il nous a semblé pertinent de dresser un tableau des taux de cohérence interindexeurs pour les différents sous-ensembles de descripteurs afin de faire ressortir les écarts de pourcentage, qui lors des analyses des protocoles, pourront guider nos analyses pour l'identification des causes du manque de cohérence.

Tableau IX. Taux de cohérence interindexeurs pour les quatre documents en fonction de divers ensembles de descripteurs.

	Taux de cohérence interindexeurs			
	<i>Étang aéré</i>	<i>Biogaz</i>	<i>Compostage</i>	<i>Énergie</i>
<b>Ensembles de descripteurs</b>				
Tous les descripteurs	47 %	28 %	40 %	44 %
DE1	68 %	27 %	40 %	51 %
DE2	2 %	7 %	----	5 %
DE géographiques	55 %	75 %	50 %	65 %
Candidats descripteurs	----	17 %	----	----

Premièrement, on remarque que les indexeurs sont plus en accord pour l'attribution des descripteurs principaux (DE1) que pour les descripteurs secondaires (DE2). En effet, la cohérence interindexeurs s'élève à 68 % pour les DE1 du document *Étang aéré* tandis qu'elle atteint un maximum de 7 % pour les DE2 du document *Biogaz*. Ces résultats viennent appuyer ce que les recherches en SI ont démontré à plusieurs reprises, c'est-à-dire que les indexeurs arrivent aisément à déterminer les quelques termes centraux d'un document (Leonard, 1977), mais que l'évaluation des notions plus secondaires (essentielle toutefois pour une représentation complète du contenu du document) requiert une plus grande part du jugement de l'indexeur, le choix des termes étant alors plus sujet à des variations interindexeurs.

Par ailleurs, les contraintes imposées par les langages documentaires et l'étendue de leur vocabulaire doivent être considérées dans les variations d'accord interindexeurs. Il a été mentionné plus tôt que les termes centraux des documents *Biogaz* et *Compostage* ne se

trouvaient pas dans les thésaurus. De fait, les taux de cohérence les plus bas reviennent à ces deux documents, soit 28 % pour le document *Biogaz* et 40 % pour le document *Compostage*. Il est donc possible que les indexeurs aient fait des compromis pour traduire leur représentation, entraînant ainsi de plus grandes variations dans le choix de descripteurs.

Les taux de cohérence les plus élevés appartiennent au sous-ensemble DEgé0 qui varie de 50 % à 75 %. Il semble vraisemblable de retrouver un plus grand accord pour ces termes puisqu'ils ne nécessitent pas de traduction et ne posent pas de problème d'interprétation.

L'attribution de candidats descripteurs étant facultative, il s'ensuit que le taux de cohérence ne peut être calculé que pour le document *Biogaz*. Soulignons que pour les documents *Étang aéré*, *Compostage* et *Énergie*, un seul indexeur a inscrit des termes dans ce champ, ce qui rend tout calcul impraticable également.

En somme, il semble que nos premiers résultats s'apparentent à ce que l'on retrouve dans les écrits sur la question.

#### 1.4 Listes de descripteurs<sup>3</sup>

Examinons maintenant les listes de descripteurs produites par les indexeurs pour les quatre documents et relevons ce qui, dans un premier temps, semble constituer les sources de variations.

Un premier examen des listes révèle une différence marquée en rapport à l'exhaustivité de la représentation des documents.

---

<sup>3</sup> L'annexe A est détachable. Elle reprend les Tableaux X, XI, XII, XIII, soit les listes de descripteurs produites par les indexeurs. Ceci rendra plus agréable la lecture en évitant les nombreux allers-retours puisque le texte fait abondamment référence aux contenus de ces tableaux.

Tableau X. Listes de descripteurs produites pour le document *Étude de sites d'implantation d'étangs aérés*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (10)</b>	<b>Tous les champs (12)</b>	<b>Tous les champs (5)</b>	<b>Tous les champs (21)</b>
<b>Descripteurs (10)</b> <b>Principaux (2)</b> étang aéré traitement eau usée	<b>Descripteurs (12)</b> <b>Principaux (5)</b> emplacement étang aéré traitement eau usée municipalité évaluation	<b>Descripteurs (5)</b> <b>Principaux (3)</b> évaluation emplacement étang aéré	<b>Descripteurs (18)</b> <b>Principaux (4)</b> étang aéré emplacement évaluation traitement eau usée
<b>Secondaires</b>	<b>Secondaires (3)</b> répercussions environnementales station épuration coût	<b>Secondaires (1)</b> traitement eau usée	<b>Secondaires (10)</b> <i>usine épuration</i> puits contamination habitation topographie superficie zonage coût dommage vent
<b>Descripteurs Géo. (8)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord Lufty Méandre Boysfarm Monfette	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord	<b>Descripteurs Géo. (1)</b> Piedmont	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord
<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs (3)</b> expropriation zone inondable design

*Note.* Un terme en italique indique une erreur de la part de l'indexeur. Soit qu'il s'agit d'un terme non préférentiel, c'est-à-dire d'un terme inscrit dans le thésaurus qui n'est pas un descripteur, ou soit que le terme inscrit n'est pas dans le thésaurus. Ce terme n'est donc pas pris en compte dans le calcul de la cohérence interindexeurs.

Tableau XI. Listes de descripteurs produites pour le document *Élimination du biogaz. Sélection d'une technologie et évaluation des impacts*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (11)</b>	<b>Tous les champs (9)</b>	<b>Tous les champs (8)</b>	<b>Tous les champs (22)</b>
<b>Descripteurs (7)</b> <b>Principaux (3)</b> élimination déchet enfouissement sanitaire pollution bruit	<b>Descripteurs (8)</b> <b>Principaux (5)</b> élimination déchet pollution air technologie répercussions environnementales gaz carbonique	<b>Descripteurs (7)</b> <b>Principaux (4)</b> technologie élimination répercussions environnementales enfouissement sanitaire	<b>Descripteurs (18)</b> <b>Principaux (5)</b> dépotoir élimination gaz technologie étude impact
<b>Secondaires (1)</b> informatique	<b>Secondaires (2)</b> <i>gestion déchet</i> échantillonnage	<b>Secondaires (2)</b> qualité air bruit	<b>Secondaires (12)</b> enfouissement sanitaire incinération qualité air analyse chimique échantillonnage effet serre odeur bruit circulation avion législation règlement
<b>Descripteurs Géo. (3)</b> carrière Miron CUM Montréal	<b>Descripteurs Géo. (1)</b> ville Montréal	<b>Descripteurs Géo. (1)</b> Montréal	<b>Descripteurs Géo. (1)</b> Montréal
<b>Candidats descripteurs (4)</b> biogaz ISCST logiciel torchère	<b>Candidats descripteurs (1)</b> caractérisation	<b>Candidats descripteurs (1)</b> biogaz	<b>Candidats descripteurs (4)</b> biogaz cogénération méthane torchère flamme visible

*Note.* Un terme en italique indique une erreur de la part de l'indexeur. Soit qu'il s'agit d'un terme non préférentiel, c'est-à-dire d'un terme inscrit dans le thésaurus qui n'est pas un descripteur, ou soit que le terme inscrit n'est pas dans le thésaurus. Ce terme n'est donc pas pris en compte dans le calcul de la cohérence interindexeurs.

Tableau XII. Listes des descripteurs produites pour le document *Compostage des boues résiduaires urbaines*.

EXPERT 1		EXPERT 2		EXPERT 3		EXPERT 4	
<b>Tous les champs (6)</b>		<b>Tous les champs (6)</b>		<b>Tous les champs (7)</b>		<b>Tous les champs (8)</b>	
<b>Descripteurs (6)</b>	<b>Descripteurs (6)</b>	<b>Descripteurs (6)</b>	<b>Descripteurs (6)</b>	<b>Descripteurs (6)</b>	<b>Descripteurs (6)</b>	<b>Descripteurs (8)</b>	<b>Descripteurs (8)</b>
<b>Principaux (5)</b>	<b>Principaux (3)</b>	<b>Principaux (3)</b>	<b>Principaux (5)</b>	<b>Principaux (5)</b>	<b>Principaux (6)</b>	<b>Principaux (6)</b>	<b>Principaux (6)</b>
épuration des eaux usées, système d' déchets liquides	composts et engrais naturels zones urbaines	composts et engrais naturels zones urbaines	composts et engrais naturels zone urbaine	déchets liquides composts et engrais naturels	déchets liquides composts et engrais naturels	technique de conversion des déchets récupération des déchets	usine de traitement des eaux d'égout technique de conversion des déchets composts et engrais naturels
composts et engrais naturels agricoles, déchets coûts/bénéfices, analyse	système d'épuration des eaux usées	système d'épuration des eaux usées	système d'épuration des eaux usées biodégradation	zone urbaine système d'épuration des eaux usées biodégradation	zone urbaine système d'épuration des eaux usées biodégradation	amélioration du sol recyclage	amélioration du sol recyclage
<b>Secondaires</b>	<b>Secondaires (2)</b>	<b>Secondaires (2)</b>	<b>Secondaires</b>	<b>Secondaires</b>	<b>Secondaires</b>	<b>Secondaires (2)</b>	<b>Secondaires (2)</b>
	techniques de conversion des déchets économique, développement	techniques de conversion des déchets économique, développement				terres agricoles déchets liquides	terres agricoles déchets liquides
<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo. (1)</b>	<b>Descripteurs Géo.</b>	<b>Descripteurs Géo.</b>
France	France	France	France	France	France		
<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs (1)</b>	<b>Candidats descripteurs (1)</b>	<b>Candidats descripteurs (1)</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>
			compostage	compostage	compostage		

Tableau XIII. Listes des descripteurs produites pour le document *Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (9)</b>	<b>Tous les champs (12)</b>	<b>Tous les champs (6)</b>	<b>Tous les champs (22)</b>
<b>Descripteurs (8)</b> <b>Principaux (5)</b> énergie solaire énergie de la biomasse énergie éolienne pays en développement ressources renouvelables	<b>Descripteurs (12)</b> <b>Principaux (5)</b> énergie éolienne énergie solaire énergie de la biomasse pays en développement utilisation des ressources locales	<b>Descripteurs (6)</b> <b>Principaux (4)</b> énergie solaire énergie éolienne bouse comme combustible pays en développement	<b>Descripteurs (22)</b> <b>Principaux (12)</b> climatologie désertification zones arides sources d'énergie non-polluante ressources renouvelables énergie solaire énergie éolienne bouse comme combustible puits <i>approvisionnement en eau</i> zones rurales énergie électrique <b>Secondaires (8)</b> précipitations vents changement climatique sécheresse utilisation des ressources locales éclairage biogaz facteurs socio-économiques
<b>Secondaires</b>	<b>Secondaires (5)</b> approvisionnement en eau dans les zones rurales éclairage biogaz <i>fumier</i> sources d'énergies	<b>Secondaires (1)</b> source d'énergie non-polluante	
<b>Descripteurs Géo. (3)</b> Sénégal Ndia Goureye Sahel	<b>Descripteurs Géo. (2)</b> Ndia Goureye Sénégal	<b>Descripteurs Géo. (1)</b> Sénégal	<b>Descripteurs Géo. (2)</b> Sénégal Sahel
<b>Candidats descripteurs (1)</b> rapport de recherche	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>

*Note.* Un terme en italique indique une erreur de la part de l'indexeur. Soit qu'il s'agit d'un terme non préférentiel, c'est-à-dire d'un terme inscrit dans le thésaurus qui n'est pas un descripteur, ou soit que le terme inscrit n'est pas dans le thésaurus. Ce terme n'est donc pas pris en compte dans le calcul de la cohérence interindexeurs.

Par exemple, pour le document *Étang aéré* (Tableau X), E3 a attribué un total de 5 DE tandis que E4 en a retenu 21 pour l'ensemble des champs de descripteurs. On remarque également que la liste de E1 ne comporte aucun DE2 comparativement à celle de E4 qui contient 10 DE2. De même, E3 n'a retenu qu'un seul DEgéo comparativement à 8 pour E1. Il en va de même pour les listes produites pour le document *Biogaz* (Tableau XI), le nombre de DE variant entre 5 et 21, et pour le document *Énergie* (Tableau XIII) dont le nombre de DE attribués varie entre 6 et 22. Pour le document *Compostage* (Tableau XII), les indexeurs inscrivent un nombre similaire de DE, soit entre 6 et 8.

Les différences les plus notables en terme du nombre de DE ressortent par la comparaison des champs de DE2, de candidats et de DEgéo. En effet, un écart important du nombre de DE2 peut être observé pour les listes des documents *Étang aéré* (0 à 10), *Biogaz* (1 à 12) et *Énergie* (0 à 8). Pour l'ensemble des documents, le nombre de candidats descripteurs varie de zéro à 4 et de 0 à 8 pour le nombre de DEgéo.

Une autre façon de repérer les dissemblances et les ressemblances consiste à observer les termes même qui composent les listes de descripteurs.

Débutons par les indexats du document *Étang aéré*. Au total, un éventail de 28 DE différents, incluant 8 descripteurs géographiques et 3 candidats descripteurs, a été utilisé pour représenter le contenu de ce document. Le taux de cohérence interindexeurs est de 47 %, soit le taux le plus élevé pour l'indexation des quatre documents.

On observe que 3 DE, *étang aéré*, *traitement eau usée*, *Piedmont*, sont utilisés par tous les indexeurs et que trois indexeurs (E1, E2 et E4) ont en commun 6 DE, soit *étang aéré*, *traitement eau usée* et *Piedmont*, *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse* et *Rivière du Nord*. De plus, E2, E3 et E4 ont tous inscrit les 5 DE suivants: *emplacement*, *étang aéré*, *traitement eau usée*, *évaluation* et *Piedmont*.

Pour l'indexation du document *Biogaz*, les listes des quatre indexeurs n'ont en commun qu'un seul DE, le DEgéo *Montréal*. Trois indexeurs (E1, E3 et E4) ont choisi le DE *enfouissement sanitaire* et le candidat *biogaz*.

Pour ce document, les indexeurs ne sont en accord que pour un seul terme, *Montréal*. Par contre, on remarque un recoupement de sens pour quelques DE (*élimination* et *élimination déchet*, *pollution bruit* et *bruit*, *pollution air* et *qualité air*) qui peut laisser sous-entendre un plus grand accord au niveau de la représentation du contenu du document, qui n'a cependant pas été traduite de façon identique lors du choix des DE.

Soulignons que cette indexation regroupe le plus grand ensemble de DE différents assignés pour représenter un document, soit 35 DE en tout. Ceci contribue à expliquer le plus faible taux de cohérence interindexeurs obtenu (28 %) pour les quatre documents.

Parmi l'ensemble des 16 DE attribués pour l'indexation du document *Compostage*, un seul DE, *composts et engrais naturels*, est retenu par tous les indexeurs, et les 3 DE *systèmes d'épuration des eaux usées*, *déchets liquides* et *France* font partie des listes de trois indexeurs (E1, E2 et E3). Un seul candidat, *compostage*, a été retenu. Le taux de cohérence pour ce document atteint 40 %.

Pour le document *Énergie*, les quatre indexeurs sont en accord pour 3 DE, *énergie solaire*, *énergie éolienne*, *Sénégal*, et trois indexeurs (E1, E2 et E3) retiennent également le DE *pays en développement*. Un seul candidat descripteur est retenu, *rapport de recherche*. Le taux de cohérence interindexeurs est de 44 %, le deuxième taux le plus élevé pour les quatre documents.

Pour l'ensemble des listes de DE produites pour les quatre documents, les indexeurs semblent d'accord sur quelques termes centraux, mais semblent cependant différer pour le choix des notions plus secondaires<sup>4</sup>. Par l'analyse des protocoles, nous tenterons de fournir

---

<sup>4</sup> Notons que les notions plus secondaires n'équivalent pas nécessairement aux DE2.

les explications à ces écarts importants de représentations, autant au niveau de l'exhaustivité de la représentation qu'au niveau de la variété des DE retenus.

### 1.5 Caractérisation des indexeurs

À partir des premières observations des listes de descripteurs, quelques points permettent de distinguer les indexeurs. Évidemment à ce stade-ci des analyses, il ne s'agit pas d'inférer sur les stratégies personnelles de chaque indexeur, mais simplement de faire ressortir les traits distinctifs de chacun en se basant uniquement sur l'observation des listes produites.

#### E1

À première vue, les indexats de E1 comportent deux caractéristiques distinctives. Un seul DE2 est retenu pour l'ensemble de ses listes et le nombre de DEgéo est nettement supérieur au nombre retenu par les autres indexeurs: 8 pour le document *Étang aéré*, 3 pour le document *Biogaz*, un pour le document *Compostage* et 3 pour le document *Énergie*. En effet, sur un total de 34 DEgéo retenus par les quatre indexeurs, près de la moitié des DEgéo proviennent des listes de E1, soit 15 DEgéo.

#### E2

Toutes les listes produites par E2 contiennent DE1, DE2 et DEgéo. Un seul candidat est utilisé. La seule observation de ses indexats ne révèlent *a priori* aucun trait majeur permettant de caractériser l'indexeur, si ce n'est que E2 est le deuxième indexeur le plus exhaustif du groupe.

#### E3

E3 peut être caractérisé comme l'indexeur le plus concis pour la représentation du contenu d'un document. En effet, alors que les indexeurs E1, E2 et E4 assignent respectivement une moyenne par document de 9, 9.75 et 18.25 DE (incluant tous les

champs), E3 ne retient que 6.5 DE par document. Peu de DE2 sont assignés, un DEgéo pour chaque document, E3 semble privilégier les DE1 pour représenter les documents. À cet égard, les indexats de E3 et E1 sont comparables.

#### E4

Avec une moyenne de 18 DE par document, E4 se distingue du groupe en tant qu'indexeur exhaustif. De fait, à l'exception du document *Compostage*, pour lequel elle n'a retenu que 8 DE, tous ses autres indexats comportent une vingtaine de DE. Cependant, le nombre de DE1 retenus par E4 est comparable au nombre de DE1 choisis par les autres indexeurs. C'est principalement au niveau des DE2 que la différence est notable. E4 retient également plus de candidats descripteurs que les autres indexeurs.

Une dernière comparaison interindexeurs en termes quantitatifs présente le taux de cohérence de chaque indexeur par rapport au groupe (Tableau XIV). En d'autres termes, ceci permet d'indiquer la proportion de descripteurs sélectionnés par un indexeur qui sont également choisis par l'ensemble des indexeurs, et ainsi, situer le degré de cohérence d'un indexeur par rapport aux taux de cohérence interindexeurs obtenus pour les quatre indexeurs pour chaque document, tel qu'indiqué au Tableau VIII.

Tableau XIV. Taux de cohérence de chaque indexeur par rapport au groupe pour l'indexation des quatre documents<sup>5</sup>.

Indexeurs	Moyenne par indexeur	DOCUMENTS			
		Étang aéré	Biogaz	Compostage	Énergie
E1	38 %	44 %	<i>23 %</i>	41 %	47 %
E2	43 %	<b>55 %</b>	24 %	<b>46 %</b>	47 %
E3	42 %	45 %	<b>37 %</b>	44 %	44 %
E4	33 %	<i>43 %</i>	28 %	<i>27 %</i>	<i>36 %</i>

Note. Pourcentage en gras = taux le plus élevé. Pourcentage en italique = taux le plus faible.

<sup>5</sup> Rappelons que le calcul du taux de cohérence de chaque indexeur par rapport au groupe est établi à partir de la moyenne de la cohérence obtenue pour toutes les paires d'indexeurs impliquant un indexeur donné (voir chapitre 2 p. 130).

E1 est l'indexeur qui s'avère être le moins en accord avec le groupe pour l'indexation du document *Biogaz*. De fait, l'observation du contenu de ses listes produites pour ce document révèle qu'il est le seul indexeur à retenir le DE *informatique*, les DEgéo *carrière Miron* et *CUM*, et les candidats *ISCST*, *logiciel* et *torchère*. Cependant, E1 obtient le taux le plus élevé (47 %) conjointement à E2 pour l'indexation du document *Énergie*.

À l'opposé, E2 est l'indexeur le plus en accord avec le groupe pour trois des quatre documents (*Étang aéré*, *Compostage*, *Énergie*). Pour l'indexation du document *Biogaz*, c'est E3 qui obtient l'accord le plus élevé.

D'un point de vue quantitatif, il semble normal que E2 soit l'indexeur le plus cohérent car il se situe entre E3 qui attribue peu de DE et E4 dont les représentations sont exhaustives. Il reste alors à vérifier si du point de vue des stratégies d'indexation, E2 serait l'agent médiateur entre des tendances divergentes pour la prise en compte des éléments de connaissance.

Le nombre élevé de DE attribués par E4 pour chaque document est probablement à l'origine de son faible accord par rapport au groupe. En effet, E4 obtient le taux de cohérence le plus faible pour l'indexation de trois documents (*Étang aéré*, *Compostage*, *Énergie*); le taux obtenu pour l'indexation du document *Compostage* n'atteint que 27 % alors que les autres indexeurs ont une cohérence par rapport au groupe se situant entre 41 % et 46 %.

En somme, les conclusions que l'on peut émettre à partir de ces premières observations prennent la forme d'un questionnement et ne nous permettent pas à ce moment-ci d'expliquer le processus d'indexation particulier de chaque indexeur et les raisons sous-jacentes aux choix différents de descripteurs.

Le plus grand accord pour le choix des DE1 comparé à celui des DE2 peut-il être associé à une utilisation d'éléments de connaissance similaires pour le choix des DE1 et à une prise en compte d'éléments différents pour le choix des DE2? Par exemple, les DE1 rendent-ils compte uniquement de la macrostructure du plus haut niveau (par exemple, le

titre), ce qui correspondrait à une stratégie d'indexation partagée par l'ensemble des indexeurs?

E1 et E3 sont les deux indexeurs les plus concis du groupe pour l'attribution de DE. Est-ce que l'apparente similitude de E1 et E3 face à une utilisation modérée de DE2 sous-tend une préoccupation commune pour certains éléments de connaissance? Ces derniers seraient-ils moins utilisés par les autres indexeurs? De même, quelle stratégie et quel(s) élément(s) de connaissance motivent une telle exhaustivité chez E4? E1 semble privilégier la représentation de la dimension géographique, mais pour quelle(s) raison(s)?

En fait, ces observations constituent autant de pistes à examiner lors des analyses des protocoles puisqu'il est difficile d'interpréter ces résultats hors de leur contexte de production. Ainsi, nous chercherons à identifier les sources des ressemblances et dissemblances de la structure et du contenu des listes afin de découvrir de quelle manière elles sont le reflet de similarités et de différences au niveau des éléments de connaissance et des stratégies utilisés par les indexeurs.

## **2. Analyses qualitatives**

La dimension qualitative des analyses repose sur les données issues des protocoles verbaux concomitants et consécutifs ainsi que des évaluations comparées. L'objectif de ces analyses consiste à mettre à jour la stratégie d'indexation propre à chaque indexeur par l'examen de l'activité d'indexation et des processus cognitifs sous-jacents afin d'expliquer les ressemblances et dissemblances de choix de descripteurs.

Cette section débute par une analyse relativement macroscopique du déroulement du processus d'indexation de chaque indexeur. Ainsi, nous identifions l'ordre chronologique suivi par chaque indexeur pour accomplir les différentes étapes de l'indexation, les diverses parties consultées du document et leurs fonctions ainsi que la nature des consultations des langages documentaires.

Par la suite, nous présentons les résultats des analyses plus fines centrées sur l'utilisation des éléments de connaissance. La première partie de ces analyses expose la manière dont chaque indexeur utilise les divers éléments de connaissance et ce que la comparaison de ces prises en compte permet d'expliquer au niveau des variations interindexeurs. La deuxième partie de ces analyses démontre l'influence prédominante de certains éléments ou combinaison d'éléments sur l'évaluation de choix de termes et descripteurs dans le processus de chaque indexeur. Finalement, nous exposons ce que cette hiérarchisation permet d'expliquer au niveau des variations interindexeurs et comment elle définit les conditions d'une indexation satisfaisante propres à chaque indexeur.

## 2.1 Déroulement de la procédure d'indexation

Sans faire de distinction quant à l'étendue de la consultation des différentes parties du document, les quatre indexeurs suivent tous la procédure d'indexation prescrite par la norme et respectent la politique du MENVIQ pour l'identification des lieux géographiques. Ils se distinguent toutefois par le moment de consultation du thésaurus et la nature de ces consultations. Ceci pointe sur leur différence de connaissance des langages documentaires et l'utilisation particulière de cette connaissance dans leur stratégie d'indexation.

Ainsi, tous les indexeurs débutent leur analyse par la lecture d'une section du document, que cette section soit choisie à l'avance ou encore, parcourue selon son ordre d'apparition dans le document. Tout au long de cette prise de connaissance du contenu, les indexeurs identifient les notions traitées et retiennent celles qui leur apparaissent comme représentatives du contenu du document<sup>1</sup> tel qu'illustré par les extraits suivants<sup>2</sup> :

<sup>1</sup> Rappelons que dans les verbalisations:

- le texte souligné représente une activité de lecture;
- le texte en **gras** représente une **activité d'écriture**;
- le texte normal se rapporte à une activité de traitement verbalisée par l'indexeur;
- le texte en *italique* se rapporte à un terme ou une notion traités, et l'italique est également utilisé pour apporter une précision pour le lecteur. Ces précisions sont généralement entre parenthèses.
- v1 et v2 se rapportent aux types de verbalisation, soit v1 pour le premier type de verbalisations recueillies, les verbalisations concomitantes; et v2 pour le deuxième type de verbalisations, les verbalisations consécutives sur trace.

<sup>2</sup> Les analyses qualitatives sont appuyées par des extraits de protocoles. Toutefois, un condensé de l'indexation d'un document, incluant verbalisations concomitante (v1) et consécutive (v2), est présenté au Tableau XV à la page 169. Si le lecteur souhaite consulter le protocole complet d'une indexation, il peut

**E1 Étang aéré v1**

(parcours de l'introduction) ...d'identifier les sites alternatifs de l'usine d'épuration. **Usine d'épuration.**

**E2 Compostage v1**

(lecture du titre) Compostage des boues résiduaires urbaines. Bon! Sujet, **compost.** La page titre là, c'est toujours mieux! Boues! Des boues résiduaires urbaines. Boues. **Boues.** OK. Je ne sais pas s'il y a plusieurs sources de boues?

**E3 Énergie v1**

(lecture du résumé) Une cueillette et une analyse approfondie est faite de données existantes reliées à ces trois sources énergétiques. Présentation de l'équipement proposé ainsi qu'un scénario... **L'équipement.**

**E4 Biogaz v1**

Je vais aller à l'introduction... Contexte. L'enfouissement des déchets au Centre de Tri et d'Élimination des Déchets a débuté en '68 par la compagnie Miron qui exploitait alors la carrière... **Enfouissement déchets.**

À cette première étape, certains indexeurs se réfèrent à leur connaissance du vocabulaire du thésaurus pour extraire des termes qu'ils savent être des descripteurs ou qui peuvent être facilement transposés en descripteurs. Ceci s'applique nécessairement à l'indexation des documents indexés avec le thésaurus familier aux indexeurs. Il s'agit en fait d'effectuer conjointement les étapes d'extraction et de traduction, procédure d'indexation potentielle soulignée par la norme internationale ISO (1985).

Cette procédure semble particulièrement privilégiée par E3 dont les listes de termes brouillon<sup>3</sup> pour les documents indexés avec Envirodoq sont constituées principalement de descripteurs.

**E3 Étang aéré v2**

(E3 inscrit emplacement sur sa feuille brouillon pour représenter le terme site d'implantation) Tant qu'à en écrire un (*descripteur*), aussi bien mettre tout de suite celui qui est dans le thésaurus. Et j'étais pas mal sûre que c'était *emplacement* de toute façon à 99 %.

D'ailleurs, E3 mentionne fréquemment de manière explicite la correspondance ou la non correspondance du terme et du descripteur.

**E3 Étang aéré v1**

Surtout *étang aéré*, je vois que c'est vraiment dans le titre, et je sais qu'en plus on l'a dans le thésaurus.

**E3 Étang aéré v1**

De toute façon *design* n'est pas dans notre thésaurus.

---

consulter l'Annexe C qui comprend la version intégrale de l'indexation du document *Étang aéré* effectué par E3 ainsi que les verbalisations consécutives.

<sup>3</sup> Rappelons que toutes les feuilles de travail, ou feuilles brouillon, utilisées par les indexeurs pour produire leurs listes finales ont été retranscrites et peuvent être consultées à l'Annexe B, également détachable.

Également, plusieurs indices dans la démarche de E3 laisse entrevoir son processus de référence constante au vocabulaire d'Envirodoq, même lorsqu'elle ne mentionne pas explicitement sa connaissance du descripteur approprié, tel qu'illustré dans l'énoncé suivant:

**E3 Biogaz v1**

Centre de tri et d'élimination des déchets. Élimination du biogaz. Sélection d'une technologie et évaluation des impacts. Mm! Fallait bien que ce soit une étude d'impact... **Biogaz. Technologie. Impact Répercussions...** (*E3 inscrit Im pour impact, le raye tout de suite et écrit répercussions environnementales qui est dans le thésaurus alors qu'impact n'y est pas*).

En effet, pour le document *Étang aéré*, cinq des huit termes inscrits sur sa liste brouillon sont des descripteurs. Les trois autres termes se rapportent à la dimension géographique. Pour le document *Biogaz*, sept termes sur neuf appartiennent au thésaurus, les deux autres étant considérés comme des candidats descripteurs potentiels puisque E3 sait qu'ils ne sont pas dans le thésaurus.

En moindre proportion, E2 et E4 reconnaissent et sélectionnent également des termes qui appartiennent au thésaurus dès la phase d'extraction.

**E2 Biogaz v1**

Évaluation des impacts. Impact. On a **répercussions environnementales** (*ce terme est le DE employé pour représenter la notion d'impacts environnementaux*), mais impact, on a quelque chose là-dessus, c'est plutôt *étude impact* (*qui est également un DE du thésaurus*).

**E4 Étang aéré v1**

Contraintes techniques. Traitement retenu. Temps de résidence en étang, respect des exigences environnementales. Temps de résidence en étang. Contraintes techniques. Là, aussi, je trouve que c'est... Je vais inscrire **contraintes techniques**. Contraintes monétaires, **coûts** (*E2 inscrit directement le DE coût plutôt que le terme contraintes monétaires*).

**E4 Étang aéré v2**

Tu vois, là ça me vient. Je sais que *coût* ça existe dans le thésaurus.

**E4 Biogaz v1**

Comparer les émissions et les retombées atmosphériques prévues avec le cadre juridique en vigueur ainsi qu'avec les plus récentes exigences canadiennes et californiennes. Ah, là, y aurait peut-être **législation**?

**E4 Biogaz v2**

Pour couvrir l'aspect de normes, règlements, lois en vigueur au Canada, un peu partout. Le terme c'est **législation**. Je savais qu'il était dans le thésaurus.

Ainsi, 4 termes des 11 retenus par E2 pour représenter le contenu du document *Étang aéré* sont des descripteurs. Trois autres termes de cette liste représentent les auteurs et un descripteur géographique. Par ailleurs, E2 sait que les termes retenus *biofiltration* et *usine épuration* ne sont pas des descripteurs. Pour le document *Biogaz*, E2 retient 6 termes qui

correspondent à des descripteurs sur les 10 termes de sa liste brouillon. Quant à E4, elle retient 12 termes qui correspondent à un descripteur sur un ensemble de 27 termes pour le document *Étang aéré*. Pour le document *Biogaz*, 9 des 25 termes de sa liste initiale sont des descripteurs. Les listes brouillon de E1 comportent 3 descripteurs sur 13 termes pour le document *Étang aéré* et 4 descripteurs sur un ensemble de 22 termes pour le document *Biogaz*, sans que E1 fasse de mention explicite de sa référence à sa connaissance du vocabulaire Envirodoq.

Cette correspondance entre termes et descripteurs ne témoigne que partiellement de la connaissance du thésaurus de chaque indexeur. De fait, ce ne sont pas tous les indexeurs qui se réfèrent explicitement à leur connaissance du vocabulaire pour la sélection des termes. Ainsi, l'indexeur peut retenir des termes sans mentionner s'il s'agit d'un terme ayant son équivalent exact dans le thésaurus, si ce terme est facilement transposable en descripteur ou s'il n'existe pas de descripteur approprié. Parfois, les indexeurs évaluent plusieurs termes qui n'apparaissent sur aucune liste, lesquels peuvent correspondre à des descripteurs. En conséquence, il devient difficile de juger de la connaissance du langage documentaire des indexeurs à partir de la seule comparaison entre les termes retenus sur les feuilles de travail et ceux des listes finales de descripteurs. De plus, soulignons que certains indexeurs consultent le thésaurus dès l'étape d'extraction.

Afin de repérer le recours effectif à leur connaissance du vocabulaire dès la première phase de l'indexation, l'on doit vérifier à quel moment du processus l'indexeur se réfère au thésaurus (mentalement ou physiquement) et identifier la nature de cette utilisation. En d'autres termes, il s'agit d'identifier les objectifs poursuivis par chaque indexeur lors de l'utilisation du langage et par extension, de déterminer l'ordre chronologique réel des différentes étapes de l'indexation suivi par chaque indexeur.

Rappelons ici les trois stratégies liées à l'exploration du langage documentaire identifiées par Bertrand (1993), et présentées précédemment au contexte théorique, puisque nous observons des objectifs similaires chez nos indexeurs dans l'utilisation (physique ou mentale) du thésaurus. Ainsi, les trois objectifs répertoriés sont les suivants:

1) *Validation d'un concept*: L'indexeur vérifie si une notion est acceptée en tant que terme d'indexation. Dans ce cas, l'indexeur consulte en général uniquement le terme d'indexation et parfois explore les termes reliés à ce dernier. Il s'agit ni plus ni moins d'une stratégie permettant d'entreprendre la transposition en descripteurs des termes jugés *indexables* préalablement à l'étape de consultation du langage.

2) *Traduction ou spécification d'un concept*: Contrairement à l'objectif précédent, l'indexeur ne recherche pas un terme particulier d'indexation, mais consulte plutôt le langage afin de préciser une notion ou exprimer la relation entre plusieurs notions. L'indexeur consulte ainsi le langage à la recherche du descripteur le plus approprié pour traduire la notion retenue. Il consulte alors les descripteurs et termes reliés ainsi que les notes d'applications ou cherche des synonymes.

3) *Extraction de nouveaux concepts-termes d'indexation*: Cet objectif consiste à utiliser le langage comme aide à l'extraction des concepts du document. Le langage est utilisé pour repérer les notions importantes du document et ainsi se construire une représentation conceptuelle de l'ensemble des idées exprimées par le document. L'indexeur choisit généralement un terme d'entrée générique dans le langage afin d'avoir accès à un vaste ensemble de terme d'indexation. Nous verrons l'utilisation de cette stratégie chez un indexeur qui n'est pas familier avec le langage documentaire.

### 2.1.1 Utilisation et connaissance du thésaurus *Envirodoq*

Des quatre indexeurs, E1 et E4 consultent le thésaurus une fois leur représentation initiale complétée, effectuant ainsi les étapes de l'indexation l'une après l'autre.

#### E1

E1 fait souvent référence à sa connaissance du domaine de l'environnement, c'est-à-dire aux thèmes et aux problématiques traités dans ce domaine, mais se réfère peu explicitement à sa connaissance du thésaurus.

#### **E1 Étang aéré v1**

Les municipalités de Piedmont, St-Sauveur-des-Monts et St-Sauveur-Paroisse se sont regroupées dans le but d'effectuer un traitement commun de leurs eaux usées se déversant dans la Rivière du Nord. (...) Alors, **traitement des eaux usées**, c'est quelque chose qui revient souvent dans les études environnementales.

En effet, à part quelques rares mentions de descripteurs, E1 semble moins bien se souvenir du vocabulaire exact d'Envirodoq. Soulignons cependant que dans l'énoncé précédent, *traitement eau usée* est un descripteur.

#### **E1 Biogaz v2**

Tu veux dire *déchet* au lieu d'*élimination des déchets*. C'est parce que tu ne sais pas comment tu vas le trouver dans ton thésaurus, *déchet*, *élimination* ou s'ils ont employé (*dans le thésaurus*) un autre mot que *élimination*, tu vas le prendre. *Biogaz* c'est pareil.

E1 tente de se donner le plus d'accès possible dans le langage en retenant plusieurs termes du document reliés à la même notion.

#### **E1 Étang aéré v1**

Mais c'est bon d'en écrire plus qu'un (*terme*) justement pour se faciliter la tâche lors de la vérification (*dans le thésaurus*).

#### **E1 Biogaz v1**

Caractérisation du biogaz. Alors (*E1 regarde les termes inscrits sur sa feuille brouillon*), tantôt on avait *composition*, *destruction*. Là, selon le thésaurus, il pourrait y avoir **caractérisation** aussi.

Ainsi, ses consultations du thésaurus Envirodoq consistent essentiellement à rechercher dans le langage chacun des termes retenus initialement. De plus, elle examine le contexte des descripteurs (TG, TS, TA et NA), ce qui lui fournit parfois d'autres pistes pour trouver les descripteurs les plus appropriés.

#### **E1 Étang aéré v1**

(*Consultation du thésaurus*) Ah, OK. Étang aéré. Étang pour traitement des eaux usées, là on utilise... l'aération mécanique, étang... Là, je lis les termes relatifs aux descripteurs, pour justement justifier mon choix, voir si...

#### **E1 Étang aéré v2**

Parce que tu peux trouver un terme qui est encore plus significatif pour ton document qui se trouve dans les termes associés ou des synonymes qui te semblent important de relever.

Par ailleurs, lorsque le terme initial retenu en brouillon ne se trouve pas dans le thésaurus, E1 cherche alors sous un synonyme, auquel elle doit penser ou sous un terme plus général qui lui donne accès à un ensemble de termes susceptibles de la mener vers des descripteurs pertinents, comme l'ensemble d'extraits suivant l'illustre:

#### **E1 Étang aéré v1**

(*À l'étape de consultation du thésaurus, E1 regarde sa feuille de travail*) Qu'est-ce que j'avais d'autre? Biofiltration. Il n'est pas là (*dans le thésaurus*). Il est peut-être sous *usine*.

(*Lecture dans le thésaurus*) On a scierie, eau industrielle, Usine d'épuration. Je vais aller voir sous *eau* seulement, peut-être qu'on va me renvoyer à des... eaux usées, terme associé, élimination eau usée.

**E1 Étang aéré v2**

T'essaies de voir des synonymes à ce que tu cherches. Soit des synonymes ou soit des termes généraux. Comme je cherchais *biofiltration*. Il n'est pas là. *Usine* non plus. Bon bien, on va aller voir à *eau* tout simplement. Il peut y avoir des voir (V) et voir aussi (VA).

**E1 Étang aéré v1**

(Lors du parcours du thésaurus) J'essaie de trouver mon terme *biofiltration*. Je regarde sous eau usée, ce n'est pas là. Je regarde sous usine. C'est pas là non plus. Usine d'épuration, utilisation eau, station d'épuration. Station d'épuration. Je ne sais pas si c'est vraiment la même chose? Employé pour (EP) usine d'épuration. J'avais *usine biofiltration*. Peut-être *filtration*? Quand on connaît pas bien le thésaurus là, c'est un peu plus laborieux. Cela prend un peu plus de temps. Filtration, filtre, terme général, traitement eau. Je vais voir à traitement eau. Employé (EM) purification pour eau potable, traitement eau usée pour les eaux usées. Alors traitement eau usée, ça serait terme employé. Épuration biologique. Alors, terme associé. Terme associé étang aéré se trouve sous traitement eau usée. Alors traitement eau usée doit apparaître.

En somme, E1 se sert en partie du thésaurus comme d'un guide pour trouver les descripteurs adéquats pour sa représentation du document.

E2 et E3 se réfèrent à leur connaissance du thésaurus dès l'élaboration de leur représentation. De plus, ces indexeurs consultent quelques fois le thésaurus à la première étape d'indexation.

**E2**

Pour l'indexation du document *Étang aéré*, E2 consulte le thésaurus après l'élaboration de sa représentation du contenu. Cependant, sa connaissance du vocabulaire d'Envirodoq lui permet de sélectionner des termes dont elle connaît la correspondance dans le thésaurus et combiner l'étape de traduction en descripteurs dès ce moment.

**E2 Étang aéré v1**

Sites d'implantation d'étangs. Bon, étang, ça c'est important. Je pense qu'on a *étang aéré* même dans notre thésaurus

**E2 Étang aéré v2**

Quand je vois *usine d'épuration*, je le sais tout de suite qu'il n'est pas dans le thésaurus. Je vois *station épuration*, et il en existe d'autres. Ça fait tout partie de *l'assainissement des eaux*. En mettant ça, il fallait que je mette... Il s'agit d'épuration et non de purification. C'est suite à ce moment là que je l'ai sorti.

Pour le document *Biogaz*, E2 applique la même stratégie en combinant les étapes d'extraction et de traduction lorsque sa connaissance du vocabulaire d'Envirodoq lui permet. Cependant, lorsque E2 juge un terme important et reconnaît qu'il est ou non un descripteur, elle s'assure le plus tôt possible de la justesse de son inférence en vérifiant la présence ou l'absence de ce terme dans le thésaurus ainsi que la forme exacte acceptée.

Ceci lui permet d'établir une représentation du contenu du document qui soit plus stable ou définitive dès ce moment de l'analyse.

### E2 Biogaz v1

...évaluation des impacts. *Impact*. On a (dans le thésaurus) *répercussions environnementales*, mais *impact*, on a quelque chose là-dessus, c'est plutôt *étude impact*. Je vais quand même aller vérifier (dans le thésaurus). (Consultation du thésaurus) *Impact*. Non, on ne l'a pas. Donc, je pense que c'est *étude impact*. Je suis quasiment certaine. Mais, ici ça semble pas être vraiment une *étude d'impact*. (Consultation du thésaurus) *Étude impact*. *Répercussion environnementale*. Terme général, terme associé: *aménagement, autorisation projet, répercussion sociale, projet*. Donc, ce n'est pas vraiment une *étude d'impact* en tant que tel. Je marquerai plutôt une *répercussion environnementale*, au premier abord. Bon.

### E2 Biogaz v1

*Destruction du biogaz produit par le site d'enfouissement du centre de tri...* Donc, *site d'enfouissement* ou... On n'a pas *site d'enfouissement*, je crois qu'on a *enfouissement sanitaire*. (Consultation du thésaurus) *Enfouissement sanitaire*. OK. *Élimination déchet*, il est quelque part. Site, ça me surprendrait. Site, site d'enfouissement. *Site historique*. Non. Donc, *enfouissement sanitaire*. (Lecture du document) *Ce rapport présente la démarche suivie pour atteindre...*

La juxtaposition des deux étapes d'indexation lui évite ainsi d'avoir à retenir en brouillon plusieurs termes pour représenter un aspect dont elle n'est pas certaine de la correspondance dans le langage documentaire.

De manière générale, les consultations du thésaurus effectuées par E2 servent trois fonctions. Premièrement, E2 confirme la présence dans le thésaurus des termes retenus en brouillon qu'elle sait être des descripteurs. Deuxièmement, elle vérifie que le sens d'un descripteur correspond de manière exacte au sens du terme employé dans le document. Par extension, E2 examine les descripteurs reliés afin de choisir le descripteur le plus approprié. Cet examen lui permet également de compléter sa représentation.

### E2 Étang aéré v1

*Implantation*. Je vais vérifier si il est présent dans le thésaurus. (Consultation du thésaurus) *Implantation, végétation*. Pas du tout rapport! Je sais pas! Là, je lis. Je dis ça pas du tout rapport mais je constate que... *NA: pour les sites n'ayant jamais eu de couvert végétal. Fermière, germination, eau végétation, carrière...* *Implantation: végétation*. Non! D'après moi, ça n'a pas rapport du tout parce que c'est, je pense que c'est plutôt pour les sites, justement, qui n'ont jamais eu de végétation. Donc, c'est peut-être des sites qui recouvrent... *Couvert végétal*.

### E2 Étang aéré v2

Je lis toujours les TG et TS pour voir. Ça m'aide beaucoup pour trouver d'autres termes ou pour dire, oui cela a bien de l'allure c'est plus spécifique encore, et ça s'applique aussi. Donc je vais plutôt que de prendre le terme général, prendre le terme spécifique. C'est pour ça que je les lis.

Troisièmement, tout comme E1, mais en moindre proportion, E2 se sert quelques fois du thésaurus pour prendre connaissance des possibilités qu'il offre, auxquelles elle ne penserait pas spontanément. Notons en dernier lieu qu'elle retient plusieurs descripteurs sur

sa liste finale sans en faire la vérification dans le thésaurus, indice de sa connaissance de ce langage documentaire.

### E3

E3 consulte le thésaurus à la phase d'extraction essentiellement pour confirmer la présence ou l'absence des termes dans le thésaurus et s'assurer de la forme exacte acceptée. Mais pour E3, il ne s'agit vraiment que d'une formalité puisqu'avant de se référer au thésaurus, elle mentionne l'adéquation entre le terme retenu et le descripteur approprié.

#### **E3 Étang aéré v1**

Évaluation des sites. Site. Dans le thésaurus, je pense que c'est emplacement. (*Consultation du thésaurus*) Emplacement, emplacement, c'est ça. **Emplacement**. Et *étangs aérés* (déjà inscrit sur sa feuille brouillon), je vais le vérifier tout de suite.

#### **E3 Étang aéré v2**

Je l'avais mis étang aéré, j'étais sûre aussi à 99 % qu'il était dans le thésaurus, il reste toujours le petit 1 %. Mais tant qu'à le mettre, je savais que le document allait parler de *l'emplacement des étangs aérés* et j'étais sûre que c'est ça que j'allais garder dans les descripteurs principaux. C'est pour ça que tout de suite je les ai vérifiés. Ça va être fait, j'aurai pas besoin de les vérifier tantôt quand je vais les réinscrire au propre.

#### **E3 Biogaz v1**

Élimination déchets, je pense qu'on l'a. (*Consultation du thésaurus*) Élimination déchets... Élimination déchets. **Élimination déchets**, tiens!

#### **E3 Biogaz v2**

Je vais tout de suite dans le thésaurus parce que tant qu'à le marquer, j'aime mieux écrire le descripteur qu'on a que de traduire après.

Par ailleurs, sa connaissance du vocabulaire d'Envirodoq lui permet de retenir plusieurs descripteurs sans en faire la vérification dans le thésaurus.

#### **E3 Étang aéré v1**

Là, *traitement eaux usées*, je ne le vérifie pas parce que je sais qu'il est dans le thésaurus.

#### **E3 Biogaz v1**

(*Lecture du document*) Évaluation des impacts, préoccupation du public, qualité de l'air. Mécanismes de la dispersion des polluants. **Qualité air**, je vais mettre ça ici (*l'inscrit en DE2 sans vérification dans le thésaurus*).

#### **E3 Biogaz v1**

(*Lecture du document*) Bilan d'énergie, bilan de masse, évaluation des impacts. Qualité de l'air, contrôle du bruit. Contrôle du bruit. **Qualité de l'air, bruit**. Je vais rajouter **bruit** aussi (*ajoute bruit aux DE2 sans consulter le thésaurus*).

Plus rarement, E3 examine les termes reliés (TG, TS, TA, NA) au descripteur choisi afin de s'assurer qu'il n'existe pas un descripteur plus approprié ou plus précis.

**E3 Biogaz v1**

*Dépotoir ou enfouissement sanitaire?* (*Consultation du thésaurus*) Dépotoir, déchets solides, restauration, ordures ménagères, enfouissement sanitaire. *Dépotoir ou enfouissement sanitaire?* (*Lecture du document*) Le site d'enfouissement. ... par le site d'enfouissement du Centre de tri.... **Enfouissement sanitaire**, tiens! C'est ça.

**E3 Biogaz v2**

Non, ça ne me disait rien. *Dépotoir* ça peut être à ciel ouvert, etc. Des fois on va parler dépotoir Miron, là je trouvais que *enfouissement* était plus approprié.

**E3 Étang aéré v1**

(*Consultation du thésaurus*) Étangs, étangs, étangs aérés. Étangs pour le traitement des eaux usées. (TG) Étang eau usée, (TA) aération, traitement eau usée, bactérie aérobie...

**E3 Étang aéré v2**

Je suis allé voir les termes reliés (*au terme étang aéré*) pour voir si je ne pouvais pas avoir d'autres idées. Pour être sûre d'avoir fait le tour de la question.

En général, ce dernier type de consultation est utilisé lors de la révision du produit global, c'est-à-dire à la fin de l'indexation. D'ailleurs, les indexeurs E2 et E4 concluent généralement leur analyse par la révision de leur liste finale de descripteurs.

**E4**

Bien que E4 ne consulte pas physiquement le thésaurus lors de l'élaboration de sa représentation du contenu, elle se réfère de façon régulière à sa connaissance du thésaurus en mentionnant la présence ou l'absence des termes dans le thésaurus.

**E4 Biogaz v1**

(*Lecture du document*) Caractérisation du biogaz. Objectif, historique des campagnes d'échantillonnage, campagne de mai '89, janvier '89 et mai '88. **Échantillonnage.**

**E4 Biogaz v2**

Ça aussi c'est un terme qui est important dans Envirodoq. *Échantillonnage, mesure.* Quand j'en voyais, je le mentionne.

**E4 Biogaz v1**

(*Lecture du document*) Synthèse des campagnes d'échantillonnage et sélection des substances les plus importantes, les composés majeurs, les composés organiques volatils. Évaluation des débits de biogaz.

**Débits.**

**E4 Biogaz v2**

Je savais que c'était un descripteur. Il va y avoir une surprise c'était pour l'eau. *Débit*, il est là dans le thésaurus alors je l'écris.

**E4 Étang aéré v1**

(*Lecture du document*) Ces contraintes sont définies brièvement ci-après et une grille de comparaison est présentée, puis discutée, pour chacun des cinq sites étudiés. Contraintes environnementales. Zonage. Les zonages actuels sont identifiés à partir des plans de zonage de la municipalité de Piedmont.

**E4 Étang aéré v2**

Je vois le mot *zonage*, évidemment il me saute au visage, parce que je sais que c'est un descripteur.

Par exemple, des 27 termes de sa liste brouillon pour le document *Étang aéré*, E4 souligne explicitement que 5 d'entre eux sont des descripteurs et que 6 autres n'en sont pas, elle n'y fait pas allusion pour les autres termes. Par ailleurs, ceci démontre également que malgré une période de deux ans sans avoir travaillé avec le thésaurus Envirodoq, E4 se rappelle relativement bien le vocabulaire de ce thésaurus.

À l'étape de traduction, ses consultations du thésaurus remplissent trois fonctions. Premièrement, E4 vérifie l'acceptation d'un terme dans le langage. En fait, il s'agit de s'assurer que le terme qu'elle croit être un descripteur est représenté sous cette forme exacte dans le thésaurus. Elle recherche donc dans le thésaurus les termes retenus à l'étape précédente et les inscrit comme descripteurs lorsqu'ils se trouvent dans le thésaurus. S'ils ne font pas partie du vocabulaire accepté, E4 cherche alors sous un synonyme.

#### **E4 Étang aéré v1**

(Début de la consultation du thésaurus) En partant avec *étang aéré*. Voilà! Alors, **étang aéré** mon premier descripteur. Le document parle de l'évaluation des *sites* proposés. Alors, *Emplacement*, **emplacement**. Alors évidemment **évaluation**. Ça serait possiblement, mes descripteurs principaux, lui je l'ai écrit, lui je l'ai écrit... (E4 raye les termes écrits en descripteurs de sa feuille brouillon)

#### **E4 Étang aéré v1**

Donc, si je fais une petite révision de mes termes. (Lecture de sa feuille de travail) Étangs aérés, sites, design. Il y a *conception*. (Consulte le thésaurus pour vérifier un synonyme de *design*) *Conception*.

#### **E4 Étang aéré v2**

Je reviens avec mon *design*. *Design*, je l'ai mis en candidat, mais j'aurais aimé cela trouver quelque chose dans le thésaurus qui aurait signifié la même chose. Là il me vient *conception* tout à coup en tête.

#### **E4 Étang aéré v1**

Ben, non! Ça n'existe pas non plus. Ouais!

Deuxièmement, E4 examine les termes (TG, TS, TA, NA) présentés en relation au descripteur consulté, ce qui lui permet parfois de prendre connaissance de descripteurs auxquels elle ne pense pas spontanément mais qui peuvent s'avérer potentiellement intéressants.

#### **E4 Biogaz v1**

(Consultation du thésaurus) Débits, volume d'eau. Bon! C'est pas ça que je voulais.

#### **E4 Biogaz v2**

Je voulais que *débit* s'applique aussi aux *gaz*.

#### **E4 Biogaz v1**

(Consultation du thésaurus) Volume d'eau s'écoulant dans un cours d'eau ou une conduite pour une période donnée. Bon! Et on ne parle pas de *débit* de *gaz*! Canalisation, polluant, période de crue. Je ne peux pas utiliser ça pour les *gaz*. (Lecture de sa feuille de travail) Captage, capture, captage. Par contre, je viens de voir *canalisation*.

**E4 Biogaz v2**

Cela fait référence au début à ce que je ne comprenais pas, les espèces de tuyaux par lesquels les *biogaz* cheminaient. Ah! y aurait-il une piste pour couvrir le *captage*?

**E4 Biogaz v1**

(*Consultation du thésaurus*) Aqueduc, conduite fermée, conduite forcée, tuyau, émissaire, égout, vanne, dérivation, prise eau, rayon hydraulique... Bon. *Canalisation*, est-ce quelque chose qui a rapport à ça? Pas vraiment, je ne crois pas!

À cette étape, E4 se distingue des autres indexeurs par une recherche élaborée de synonymes pour tenter de traduire les termes retenus en brouillon qui n'ont pas d'équivalent direct dans le vocabulaire. Nous verrons plus loin que cette stratégie est motivée par son souci d'exhaustivité.

### 2.1.2 Révision de l'indexation

Cette étape n'est pas appliquée de manière systématique chez les indexeurs. Néanmoins, la révision a pour but principal de vérifier que rien n'a été oublié.

#### E1

E1 est le seul indexeur à ne pas réviser son produit final. Notons toutefois que E1 procède de cette manière pour l'indexation des quatre documents.

#### E2

Lors de sa révision, E2 revient au titre du document et évalue ses descripteurs. Elle relit également sa liste de termes pour s'assurer qu'aucun aspect n'a été oublié et que sa représentation en termes officiels est adéquate. Elle apporte parfois des modifications à sa liste finale. Il est cependant rare qu'elle consulte le thésaurus à cette étape. En fait, il s'agit pour E2 d'une évaluation globale du produit.

E3

E3 consulte l'introduction ou la table des matières lors de sa révision finale afin de s'assurer de n'avoir rien oublié. Comme il a été mentionné précédemment, E3 consulte parfois les termes reliés au descripteur retenu afin de choisir le descripteur le plus précis.

E4

Pour E4, l'étape de révision consiste principalement à comparer sa liste de termes initiaux à sa liste de descripteurs et tenter de trouver des synonymes dans le thésaurus pour les termes inscrits sur la feuille de travail qui n'ont pas été traduits en descripteurs.

L'identification de la procédure d'indexation des indexeurs a fait ressortir l'optimisation que font les indexeurs de leur connaissance du thésaurus et de la tâche d'indexation. Ainsi, le fait de retenir, de vérifier ou de s'assurer de l'adéquation des termes avec des descripteurs dès la première phase de l'indexation semble rendre compte de l'expertise des indexeurs. Bien que cette stratégie ne soit pas prescrite dans les normes, elle semble témoigner d'une adaptation aux contraintes de la tâche d'indexation. En effet, il apparaît normal que les indexeurs aient développé une approche leur permettant d'accomplir leur tâche de manière efficace, étant donné le peu de temps dont ils disposent habituellement pour analyser un document.

### *2.1.3 Utilisation du thésaurus Infoterra*

La procédure d'indexation globale ayant été sommairement exposée, nous examinons maintenant les conséquences de l'introduction d'un nouveau langage sur la procédure de chaque indexeur. Trois éléments seront observés: le moment de consultation du thésaurus, la nature des consultations et les différentes sections consultées. Concernant ce dernier point, rappelons que le thésaurus Infoterra présente son vocabulaire selon quatre catégories différentes (sommaire des catégories et des sous-catégories; liste des termes d'après leur numéro de code; liste de termes d'après leur catégorie; liste alphabétique des termes).

## E1

Pour E1, on ne constate aucun changement de procédure avec l'utilisation du thésaurus Infoterra. Ainsi, E1 procède exactement de la même façon avec les deux thésaurus, c'est-à-dire qu'elle construit sa représentation du contenu du document avant de consulter le thésaurus. Ses consultations se rapportent aux mêmes fonctions, le thésaurus servant principalement de guide pour trouver les descripteurs les plus appropriés pour traduire la représentation initiale.

**E1 Compostage v1**

(Consultation du thésaurus) Boue n'y est pas. *Épuration*, c'est système d'épuration des eaux usées, ce n'est pas vraiment ça dont on parle, c'est le traitement des eaux usées. Bon! À eau, peut-être? Eaux usées: décharge. Ah, déchets liquides. Déchets, déchets liquides, déchets industriels liquides. Eaux usées industrielles. Évacuation des eaux usées, déchets municipaux, influents industriels... *Déchets liquides*, ça me semble être la même chose que *boues résiduaire*s. Qu'est-ce qu'on disait encore sur *eaux usées*? Déchets, eaux industrielles. Déchets liquides. Mm... (*Regarde sa feuille de travail*) *Boues d'épuration* et *épuration*. *Système d'épuration des eaux usées*. C'est pas ça. Bon! **Déchets liquides**.

**E1 Énergie v1**

Il y avait *recommandation* (terme inscrit sur sa feuille de travail) (Consultation du thésaurus) Il n'est pas là non plus. C'est plate ça! *Recommandation*, c'est important, c'est pas dans toutes les études qu'il y en a. Ah! Renouvelables, ressources. Tiens, tiens! Sous ressources naturelles, ressources non renouvelables, ressources renouvelables. (*Regarde alternativement sa feuille et le thésaurus*) Oui. Énergies renouvelables, ressources renouvelables.

E1 limite cependant son exploration d'Infoterra à la liste alphabétique des termes.

## E2

E2 consulte Infoterra tôt dans son analyse et sort une première représentation du document, composée essentiellement de descripteurs appartenant à Envirodoq et complète sa représentation en consultant le vocabulaire d'Infoterra. La stratégie sous-jacente se rapporte au même but que pour l'indexation des deux premiers documents, soit d'effectuer l'étape de traduction simultanément à l'élaboration de la représentation.

**E2 Compostage v1**

Je cherche *compost* (E2 consulte le thésaurus pour ce terme déjà retenu en brouillon). Composts et engrais naturels, engrais chimiques. Donc, j'imagine qu'on peut utiliser *composts et engrais naturels*, les deux termes. Parce que ce n'est pas *engrais chimiques*. (*Retourne au document*) La maturation, l'évaluation, classification, procédé... Procédé, technologies. L'épandeur à fumier, projet de compostage (...) des boues liquides, des boues déshydratées... Je cherche si ça concerne vraiment les *engrais chimiques* ou *composts et engrais naturels*. Compostage des boues résiduaires urbaines... Je vais relire un peu la préface pour me situer là.

Ses consultations du thésaurus Infoterra sont de même nature qu'avec l'utilisation d'Envirodoq, c'est-à-dire que E2 vérifie l'adéquation du sens de terme avec celui du descripteur, elle prend connaissance des descripteurs reliés et parfois, elle se sert du vocabulaire comme aide à l'élaboration de sa représentation.

### **E2 Énergie v1**

(*Lecture du document*) Aperçu sur les milieux physique et humain. Milieu... milieu. (*Consultation du thésaurus*) Milieu marin. Milieu rural, je vois ici, employer zones rurales, et tantôt j'avais *régions rurales* (terme inscrit sur sa feuille de travail). Donc, employer zones rurales. Un mot à retenir.

### **E2 Énergie v2**

En voyant *milieu marin*, je voyais les autres milieux et je suis tombé sur *milieu rural*. Donc, ils n'appellent pas ça *milieu rural*, c'est vraiment *zone rurale*. Donc, j'en ai profité, je l'ai mis, j'ai dit, si je veux le mettre c'est sous ce terme-là.

Sa non connaissance du vocabulaire de ce thésaurus oblige E2 à consulter ce dernier plus longuement afin de trouver les descripteurs appropriés. Notons qu'elle ne peut cependant pas inscrire de descripteurs directement sans consultation du thésaurus puisque ce langage ne lui est pas familier.

E2 consulte principalement la liste alphabétique et se réfère brièvement à la section sommaire des catégories.

### **E3**

L'introduction d'un nouveau thésaurus a entraîné plusieurs changements dans la procédure de E3. Premièrement, E3 effectue les deux étapes principales de l'indexation de manière consécutive plutôt que simultanée.

### **E3 Compostage v2**

J'aimais mieux faire le tour du document, sortir ce qui me semblait important, et après aller voir dedans (*le thésaurus*). Je ne connaissais pas le thésaurus, ç'aurait été trop long à chaque terme d'aller chercher directement.

Comme pour l'indexation des deux premiers documents, elle pense également en fonction du but final de l'indexation dès le début de son analyse. Ne connaissant pas à l'avance les possibilités offertes par le langage documentaire, elle se donne plus de pistes de recherche dans le thésaurus en retenant plus de termes dès cette étape. Sa représentation

initiale est alors un peu plus exhaustive que pour l'indexation des deux documents indexés avec Envirodoq.

### **E3 Énergie v2**

Ne connaissant pas le thésaurus, je ne connais pas nécessairement les termes. Donc, je ne savais pas *énergie éolienne* comment je vais retrouver ça là-dedans. *Solaire* non plus, *biomasse*, *biogaz*, *fumier* est-ce que je vais retrouver ça là-dedans? Donc, je me donne d'autres pistes. *Méthane*, *matière organique*, *gaz carbonique*, c'est d'autres pistes pour la production de ce genre d'énergie-là, que je vais pouvoir aller chercher là-dedans, si je ne trouve pas ce que je veux.

Notons que la plupart des termes retenus en brouillon pour le document *Compostage* correspondent à des descripteurs d'Envirodoq, bien que E3 sait qu'elle devra en faire la traduction par la suite.

### **E3 Compostage v2**

J'ai sorti ce qui me semblait important. Mais j'employais des mots du vocabulaire de notre thésaurus. Et je me suis dit, je traduirai avec ça après.

Pour l'indexation du document *Énergie*, E3 adapte sa stratégie à cette nouvelle condition et sélectionne plus de termes correspondant au vocabulaire du document plutôt qu'au vocabulaire d'Envirodoq.

Évidemment, les consultations du thésaurus Infoterra effectuées par E3 ne sont pas de même nature que ses consultations d'Envirodoq. Alors que E3 ne fait que confirmer la présence des descripteurs pour l'indexation des deux premiers documents, elle procède à une exploration exhaustive du thésaurus Infoterra. Ainsi, elle explore en premier le sommaire des catégories et sous-catégories pour avoir une représentation de la segmentation du domaine établie dans ce thésaurus et le vocabulaire employé. Ensuite, elle cherche dans la liste alphabétique les termes retenus à la première étape et prend connaissance des termes apparentés en parcourant également la liste par catégorie afin de choisir le descripteur le plus adéquat pour exprimer la notion du document. Cette consultation exhaustive requiert nécessairement plus de temps pour l'étape de traduction qu'avec l'utilisation du thésaurus connu. Ainsi, pour l'indexation des deux derniers documents, E3 consulte le thésaurus une fois sa représentation extraite du document et consulte principalement deux types de listes pour repérer les descripteurs, soit alphabétique et par catégorie.

## E4

Avec Infoterra, E4 conserve la même séquence dans sa procédure, c'est-à-dire qu'elle consulte le thésaurus après avoir extrait sa représentation du contenu du document. E4 a également tendance à retenir sur sa feuille de travail des termes appartenant au vocabulaire d'Envirodoq, procédure visant à effectuer les étapes d'extraction et de traduction simultanément.

**E4 Énergie v1**

(*Lecture du document*) Recyclage de la fumure animale en biogaz devant servir à produire l'électricité nécessaire à l'éclairage. Recyclage, déchet animal.

**E4 Énergie v2**

*Recyclage, c'est toujours en rapport avec les énergies renouvelables. Recyclage de la fumure animale, donc déchet animal. Ça c'est un réflexe d'Envirodoq, parce que c'est un descripteur (déchet animal), j'ai mis celui-là.*

Ses consultations relèvent de fonctions différentes. Ainsi, plutôt que de chercher systématiquement la présence dans la liste alphabétique des termes retenus en brouillon, E4 explore la liste par catégories et sous-catégories. La consultation de la liste par catégorie permet à l'indexeur de prendre connaissance de la façon dont ce langage segmente le domaine de l'environnement et de choisir une entrée générique (soit dans la liste par numéro de code ou alphabétique) susceptible de conduire à l'ensemble des termes reliés à la notion devant être représentée.

**E4 Énergie v1**

(*Consultation du thésaurus: liste d'après leur numéro de code*) Humidité, effets microclimatiques, élévation du niveau de la mer, modifications météorologiques, bon! Lithosphère. Ça non! Écosystèmes terrestres: sols, terres agricoles, alcalines, contaminées, de bruyère, capacités optimales du sol... Ouais, peut-être ça? Pollution du sol, mise en valeur des terres, rétablissement du sol... amélioration, salination, érosion hydraulique. Terres arides et désertification: zones arides... désertification, ça je l'avais en plein!

Cependant, E4 ne semble pas avoir d'approche privilégiée pour l'exploration de ce nouveau thésaurus, alternant de la liste par catégorie à la liste par numéro de code et consultant quelques fois la liste alphabétique.

En fait, E4 éprouve quelques difficultés d'utilisation avec ce thésaurus, ce qui altère sa procédure d'exploration. E4 souligne avoir l'impression "d'aller à la pêche aux descripteurs" et d'adapter le document au thésaurus plutôt que de chercher les descripteurs qui conviennent à la représentation du contenu du document. Ainsi, dès que E4 rencontre

un descripteur qui correspond au contenu, elle l'inscrit afin d'éviter de rechercher à nouveau cette notion dans le thésaurus.

Par ailleurs, E4 attribue également à la difficulté d'utilisation de ce nouveau langage, le fait de ne pas avoir tracé une démarcation claire entre les deux niveaux de descripteurs (DE1, DE2) pour l'indexation des deux derniers documents.

**E4 Compostage v1**

Là, je me rends compte que je les ai tous mis en descripteurs principaux.

**E4 Compostage v2**

Ça je suis certaine que c'est à cause du thésaurus. J'étais tellement contente d'en avoir trouvé, pas beaucoup, six, que je les ai tous placés en descripteurs principaux.

En effet, les descripteurs représentant le coeur du document ne se retrouvent pas en début de ses listes comme pour les documents *Étang aéré* et *Biogaz*, mais plutôt entremêlés parmi l'ensemble des descripteurs. C'est-à-dire que l'ordre des descripteurs correspond à l'ordre dans lequel les termes ont été retenus à l'étape d'extraction plutôt qu'à un ordre attribué une fois les descripteurs choisis.

#### *2.1.4 Utilisation des éléments textuels (indices structuraux)*

Chaque indexeur doit parcourir le contenu du document, indépendamment du fait qu'il effectue conjointement ou séparément les deux principales étapes de l'indexation. Cette consultation des différentes parties du document permet également de caractériser de manière globale la lecture de chaque indexeur et d'identifier les différences interindexeurs à ce niveau.

Nous avons vu au contexte théorique que selon la théorie de compréhension de texte (van Dijk et Kintsch, 1983) chaque section du document remplit une fonction particulière. Par exemple, le titre rapporte habituellement le thème du document, l'introduction et le résumé présentent succinctement le développement du thème ou des thèmes abordés et la table des matières représente l'organisation hiérarchique de l'ensemble des concepts traités dans le document et indique les relations entre les concepts.

Ainsi, l'on suppose que les indexeurs se comportent comme la plupart des lecteurs et utilisent les différentes parties d'un document en accord avec leur fonction première, c'est-à-dire repérer le thème premier du document dans le titre, saisir les liens entre les concepts à partir de la table des matières, comprendre l'idée générale du document en parcourant le résumé, etc.

Concrètement, les indexeurs suivent la prescription de la norme ISO (1985) quant aux différentes parties du document à parcourir. En effet, tous les indexeurs consultent la page couverture, la page titre, la table des matières, l'introduction, le résumé (pour le document *Énergie*) ou le sommaire (pour le document *Biogaz*), ainsi que plusieurs grands titres dans le corps du document.

Tous les indexeurs extraient plusieurs termes du titre, de la table des matières et de l'introduction ou du résumé. On remarque que le titre et la table des matières sont les sections privilégiées des indexeurs pour extraire les informations essentielles et pour réviser leur représentation. L'introduction et le résumé viennent en seconde position et servent principalement à clarifier l'idée que les indexeurs se font du contenu du document et à préciser les notions rencontrées aux sections précédentes; ce qui correspond à la fonction attribuée à ces sections dans le modèle de van Dijk et Kintsch (1983). Les extraits suivants témoignent du rôle attribué à différentes parties du document par les indexeurs.

### **E3 Étang aéré v2**

Je commence avec le titre, ça me donne déjà des bons points de départ. J'ai déjà *étang aéré, site*. La table des matières aussi, je vois en gros ce qu'il va y avoir. Et là l'introduction, c'est un peu comme un titre, mais en plus détaillé avec vraiment le contenu du document. Et la conclusion c'est plus pour vérifier, pour voir si je n'ai rien oublié, et pour me confirmer dans mes choix.

### **E4 Biogaz v2**

Les descripteurs principaux que j'ai mis, je regarde le titre de mon document, paf! c'est ça. Ça parle de ça. Avec ces quatre (*élimination déchet, technologie, étude impact et biogaz*), je suis certaine que j'ai couvert le principal dans le document.

De plus, les indexeurs se réfèrent parfois à la table des matières lorsqu'ils sont indécis face à l'importance d'une notion rencontrée en introduction, dans le résumé ou à l'intérieur du document. Les énoncés suivants rendent compte de cette fonction.

### **E1 Compostage v1**

Ordures ménagères. Est-ce que ça vaut la peine d'en parler? On va retourner à la table des matières. Filières de compostage des boues déshydratées. Non! (~~*raye ordures ménagères*~~) C'est pas dans la table des matières.

**E2 Biogaz v1**

*(Lit dans le sommaire)* Toutefois, les études ont permis de conclure que la composition typique qui doit être utilisée pour la conception est la suivante: Gaz, méthane. Je suis en train de me demander si ces gaz-là, s'il va falloir les nommer. Je vais aller voir à la table des matières, s'ils en font mention. *(Vérifie dans la table des matières)* Non, il ne semble pas mentionner les gaz comme tels. Donc, je ne nommerai pas les gaz spécifiques, comme par exemple, méthane.

La conclusion est généralement utilisée pour confirmer les choix de termes de la représentation et pour s'assurer de ne rien oublier, comme le souligne l'énoncé de E1.

**E1 Étang aéré v2**

La conclusion confirme ton choix de descripteurs. Ça te permet de retracer ceux que tu n'aurais pas vu.

Ainsi, nos indexeurs font une utilisation similaire des différentes sections d'un document. Ils se distinguent toutefois par une lecture ponctuelle ou plus en profondeur de certaines sections. E2 et E3 lisent les grands titres de section et consultent l'intérieur du document pour vérifier le sens d'un concept, clarifier un aspect non compris ou vérifier que rien n'a été oublié.

**E2 Biogaz v2**

*(Dans la table des matières)* Évaluation des débits de biogaz, c'est important. Ils en parlent et il y a une bonne partie là-dessus. Donc, je devais souligner cela. C'est une partie que je devais plus approfondir pour savoir c'est quoi ça. Je devais aller dans le document voir si c'était décrit *(E2 va voir à la page 20 du document où la notion est traitée)*.

**E3 Étang aéré v2**

*(E3 feuillète le document)* Je fais souvent cela. Pour être sûre que je n'oublie rien.... Je regarde les grands titres, les caractères différents, les caractères gras. Je regarde vite, vite.

Malgré quelques consultations très ponctuelles à l'intérieur du document, E3 se limite habituellement à une lecture partielle des parties plus élaborées, qu'il s'agisse d'une préface, d'un sommaire ou d'une introduction. Tout comme E2, elle lit alors rapidement la première phrase des paragraphes afin de saisir s'il y a matière à un traitement plus en profondeur. En effet, pour ces indexeurs, le niveau de lecture d'une section est généralement déterminé par sa longueur. Un sommaire de cinq pages est parcouru rapidement, une introduction d'une demi-page est lue entièrement, etc.

En comparaison, E1 et E4 ont tendance à parcourir plus de sections (listes des tableaux, des annexes, préface, avertissement, corps du document, etc.) et à en faire une lecture plus en profondeur. Par exemple, pour l'indexation du document *Biogaz*, E1 a

consulté une dizaine de pages du corps du document en plus des sections consultées par l'ensemble des indexeurs. De même, pour l'indexation du document *Compostage*, E4 a fait une lecture intégrale de la préface, de l'introduction et de la page d'avertissement. Il est à noter toutefois que la majorité des termes retenus par E1 et E4, à la première étape de l'indexation, sont issus essentiellement des sections consultées par tous les indexeurs.

Ce parcours plus exhaustif des documents peut être à l'origine de leurs listes de termes brouillon plus exhaustives que celles des indexeurs E2 et E3. Aussi, la lecture plus en profondeur de certaines sections peut probablement être associée à certains choix de termes qui, rencontrés plus fréquemment, peuvent acquérir une plus grande importance pour ces indexeurs.

Par ailleurs, la connaissance du domaine joue peut-être un rôle dans l'examen différent du document. En effet, rappelons que E1 et E4 n'avaient pas indexé de documents appartenant au domaine de l'environnement depuis environ deux années, ce qui a pu les amener à vouloir se replacer dans le contexte de ce domaine en prenant connaissance du contenu du document par une lecture plus détaillée.

Les comportements observables des indexeurs, liés à la consultation des thésaurus et à la lecture des documents, ne permettent que des inférences fort incertaines et trop parcellaires des différences de choix de termes. En nous basant sur l'utilisation des divers éléments de connaissance, nous allons tenter de mettre à jour ce qui constitue une indexation satisfaisante pour chaque indexeur.

## 2.2 Repérage de l'utilisation des éléments de connaissance

L'hypothèse de base de notre recherche implique que les conditions du *satisficing* varieront d'un indexeur à l'autre, mais demeureront relativement constantes pour un même indexeur. Ainsi, pour pouvoir identifier et caractériser les stratégies d'indexation d'un indexeur, on doit premièrement repérer les différents éléments de connaissance utilisés par l'indexeur.

Rappelons que les éléments de connaissance auxquels nous nous attarderons se subdivisent en deux groupes: 1) les critères relatifs à la couverture sémantique du contenu du document, c'est-à-dire ceux qui font référence plus directement à l'adéquation entre les termes choisis et le contenu des documents, soit les critères de qualité de l'indexation (exhaustivité, exactitude, spécificité); et 2) les éléments relatifs au contexte de la tâche, c'est-à-dire les éléments qui font référence à l'adéquation entre les termes choisis et l'utilisation qui en sera faite, soit la prise en compte de l'utilisateur et la pertinence.

On examine alors les protocoles à la recherche de la mention et de la prise en compte des divers éléments de connaissance. Il s'agit ensuite de rassembler et d'analyser tous les énoncés se rapportant à un même élément pour caractériser l'utilisation particulière qu'en fait cet indexeur. La comparaison des résultats issus de cette analyse pour les quatre indexeurs devrait permettre d'identifier les divergences interindexeurs qui peuvent être à la source des choix différents de termes et descripteurs.

Pour que le lecteur ait un aperçu du déroulement continu d'une indexation et du type de matériel à partir duquel les analyses ont été réalisées, nous débutons cette section des analyses par un résumé d'une indexation accompagné d'indications relatives aux éléments de connaissance évoqués et à la procédure d'indexation.

L'indexation présentée est celle de E3 pour le document *Étang aéré*. Cette indexation a été complétée en douze minutes et la transcription des verbalisations concomitantes (v1) et consécutives (v2) totalisent vingt pages de texte. Pour les fins de la démonstration, nous avons réduit ce texte à quatre pages en sélectionnant principalement les extraits se rapportant aux choix et rejets de termes et descripteurs. Nous avons également retenu les énoncés permettant d'identifier l'utilisation de plusieurs éléments de connaissance ainsi que des énoncés témoignant plus explicitement d'une stratégie d'indexation.

Au Tableau XV, la colonne de gauche sert à numéroter les extraits de protocole. La colonne centrale comprend les verbalisations concomitantes ou v1 (encadrées pour la démonstration présente afin de faciliter la distinction entre v1 et v2) et les verbalisations consécutives. La colonne de droite regroupe les commentaires relatifs à l'utilisation des

divers éléments de connaissance qui servent à caractériser la méthode d'indexation personnelle de l'indexeur, ainsi que l'identification de stratégies.

Soulignons que la résolution de l'indexation étant un processus intellectuel complexe faisant appel à un vaste bassin de connaissances, plusieurs éléments peuvent être utilisés conjointement pour évaluer une action. De plus, rappelons que ce condensé doit être considéré uniquement à titre démonstratif puisqu'il ne saurait témoigner complètement de l'ensemble des stratégies d'indexation et des éléments de connaissance utilisés par E3. Ainsi, les constatations faites à partir de ce protocole seront précisées et confortées par les analyses détaillées des quatre indexations de E3 afin d'établir précisément le profil de cet indexeur.

Tableau XV. Condensé de l'indexation du document *Étang aéré* et identification partielle de la procédure d'indexation et des éléments de connaissance utilisés par E3<sup>4</sup>.

Extrait	Protocoles concomitant et consécutif	Commentaires
1	<p>(Lecture de la page couverture) Société Québécoise d'assainissement des Eaux. Programme d'assainissement des Eaux du Québec (P.A.E.Q.). <u>St-Sauveur-des-Monts, St-Sauveur-Paroisse, Piedmont. Étude de sites d'implantation d'étangs aérés.</u> <i>Étangs aérés</i> là, je vais écrire ça. <i>Étangs aérés...</i></p> <p><b>expérimentateur</b><sup>5</sup> : Et tu l'écris tout de suite?  <b>E3</b>: Oui, parce que je sais qu'il est dans le thésaurus. Donc, je l'ai écrit tout de suite pour ça. Aussitôt que je les lis, les mots qui me semblent qui pourraient être des descripteurs, je les écris tout de suite en partant.</p>	<p>Sélection des termes potentiels.  Reconnaissance d'un descripteur du thésaurus.</p>
2	<p>Donc j'écris tout ce que je pense d'un premier jet et après j'y vais un peu par élimination.</p>	<p>Méthode d'extraction de l'information.</p>
3	<p>Surtout <i>étang aéré</i>, je vois que c'est vraiment dans le titre, et je sais qu'en plus on l'a dans le thésaurus.</p>	<p>Élément textuel évoqué pour la sélection de termes : titre  Reconnaissance d'un descripteur du thésaurus.</p>
4	<p><u>Consultants. Société Québécoise d'assainissement des eaux. Étangs aérés à Piedmont. Piedmont. Tiens, Piedmont.</u></p> <p><b>expé</b>: Tu l'as écrit après avoir lu la page titre (<i>deuxième page</i>)?  <b>E3</b>: Oui, <i>Piedmont</i>, je l'ai écrit parce qu'il était là (<i>dans le titre de la page titre</i>). Je suis habituée de mettre des descripteurs géographiques, tout de suite j'ai identifié que c'était à <i>Piedmont</i>. Je l'ai écrit pour ça.</p>	<p>Terme géographique sélectionné à partir du titre.  Application de la politique d'Envirodoq.</p>

<sup>4</sup> Rappelons que la version complète de l'indexation du document *Étang aéré* effectué par E3 se trouve en Annexe C.

<sup>5</sup> Le terme expérimentateur sera dorénavant exprimé par l'abréviation expé.

5	<p><u>Étangs aérés, site d'implantation et critères de design.</u> <u>D'implantation, design des étangs.</u> <i>Étangs aérés</i>, je l'ai!</p> <p>expé: Peux-tu m'expliquer un peu qu'est-ce que tu fais? E3: Là je me suis dit, <i>critères de design</i>, on fait le <i>design</i> de quoi? Des <i>étangs aérés</i>. Donc le sujet c'est encore, ça confirme mon choix d'avoir mis <i>étang aéré</i>.</p>	Critère de sélection: La récurrence d'un terme comme sujet confirme la sélection.
6	Mais je ne mettrais pas nécessairement <i>critères de design</i> parce c'est juste un aspect qu'on étudie.	Ne retient pas les aspects plus secondaires (critère de design est représenté par ou inclus dans le terme sélectionné <i>étang aéré</i> réf. extrait 5)
7	À moins que le document porte entièrement sur le <i>critère de design</i> .	Critère de sélection: retenir thème principal, reconnu par l'ampleur du traitement.
8	<p><u>Évaluation des sites.</u> <i>Site</i>, mais je pense que dans le thésaurus, c'est <i>emplacement</i>. (vérifie dans le thésaurus) <u>Emplacement.</u> <b>Emplacement</b></p> <p>expé: Tu voulais marquer <i>site</i>? E3: Oui. Là je vois <i>évaluation des sites</i>, une grande partie du document c'est les <i>sites</i>. Je me suis dit, on va évaluer les <i>sites</i>. C'est presque tout le document (<i>en regardant dans la table des matières</i>).</p>	Critère de sélection: ampleur du traitement d'un terme.
9	Puis <i>site</i> , je sais que nous autres on ne l'a pas employé. C'est <i>emplacement</i> qu'on a dans le thésaurus. Tant qu'à en écrire un, aussi bien mettre tout de suite celui qui est dans le thésaurus.	Procédure d'indexation: exécuter deux étapes conjointement (extraction et traduction).
10	<p><u>Introduction.</u> <u>Les municipalités de Piedmont, St-Sauveur-des-Monts et St-Sauveur-Paroisse se sont regroupées dans le but d'effectuer un traitement commun de leurs eaux usées.</u> <b>Traitement eaux usées.</b></p> <p>expé: Tout de suite comme ça? E3: Oui. Si on évalue les <i>emplacements</i> pour des <i>étangs aérés</i>, ça veut dire qu'on a l'idée, c'est pour traiter des eaux usées. Le but c'est ça.</p>	Critère de sélection: but du document.
11	Donc là je pensais, si quelqu'un fait une recherche un peu plus générale sur le <i>traitement des eaux usées</i> , il va être intéressé à avoir ce document-là. C'est pour cela que je l'ai retenu.	Usager: Choix d'un terme en fonction d'une requête potentielle.
12	Mais je ne savais pas encore si je le mettais en descripteur principal ou secondaire.	Procédure d'indexation: À la première étape d'indexation, choix des termes sans attribution immédiate de statut.
13	<p><u>Dans le but d'effectuer un traitement commun de leurs eaux usées se déversant dans la Rivière du Nord, Piedmont!... À St-Sauveur-des-Monts, je vais rajouter ça. Puis St-Sauveur-Paroisse.</u></p> <p>expé: Tu reviens, tu dis <i>Piedmont</i>, tu regardes sur la page titre et là tu ajoutes. E3: Là j'ai rajouté <i>St-Sauveur-des-Monts, St-Sauveur-Paroisse</i>. J'ai pensé, <i>Rivière du Nord</i>, c'est un lieu cela, je vais sauter dans les lieux. Ici on parlait de la municipalité de <i>Piedmont, St-Sauveur-des-Monts, St-Sauveur-Paroisse</i>. Alors je les ai écrits pour ne pas les oublier, vu qu'on était dans les descripteurs géographiques.</p> <p>expé: qu'est-ce qui est arrivé avec <i>Rivière du Nord</i>? Tu</p>	<p>Attribution de DEgé : retenir terme géographique. Application de la politique d'Envirodoq.</p> <p>La récurrence de ces termes influence probablement leur choix à ce moment-ci.</p>

14	<p>t'es comme arrêté dessus...</p> <p>E3 Le fait que les eaux usées se jettent dans la <i>Rivière du Nord</i> cela a plus ou moins d'importance.... Et <i>Rivière du Nord</i>, étant donné qu'on n'en parle pas. Que les eaux usées se déversent dans la <i>Rivière du Nord</i> ou dans une autre, ce n'est pas le but, ce n'est pas le document. Je me suis dit que ce n'était pas le document.</p>	Rejet d'un terme dont le document ne traite pas suffisamment et qui ne se rapporte pas au propos central du document.
15	<p><u>L'emplacement déterminé pour les étangs aérés étant situé dans une zone à vocation résidentielle, ... a mandaté le consortium Gendron afin d'identifier des sites alternatifs d'implantation de l'usine d'épuration....</u> C'est station d'épuration. <b>Station d'épuration.</b> (<i>le DE à employer pour représenter usine d'épuration c'est station d'épuration</i>)</p> <p>expé: qu'est-ce qui se passe là? E3: ...Là, je savais que ça parlait d'<i>étang aéré</i> mais peut-être que ça pouvait aller plus loin aussi. Ça pouvait aller jusqu'à la construction de <i>l'usine</i>. J'ai pensé écrire <i>station</i>, comme ça pour ne pas l'oublier si jamais le document va plus loin là-dedans.</p>	<p>Reconnaissance d'un terme du thésaurus.</p> <p>L'ampleur du traitement évoqué comme critère de sélection.</p>
16	<p><u>Cinq différents. Bon! Chaque usine sera jugée dans le présent document en fonction de contraintes environnementales. Le présent rapport résume les avantages et les inconvénients de chacun.</u></p> <p>expé: Tu passes rapidement quand tu lis les contraintes. T'as rien traité au niveau des contraintes? E3: Non. Qu'est-ce qui est important c'est qu'on les évalue. Mais les types de <i>contraintes</i>, non je n'avais pas jugé plus que ça.</p>	Semble rejeter les aspects jugés plus secondaires (en référence au terme retenu évaluation. réf. extrait 22 )
17	<p>Puis <i>contraintes</i>, je sais qu'il n'y a rien dans le thésaurus qui va réussir à rendre cela. Donc ça règle mon problème.</p>	<p>Utilisation de sa connaissance du thésaurus.</p> <p>Semble moins s'attarder aux aspects qui ne peuvent être représentés par le vocabulaire du TH.</p>
18	<p>Si cela avait été vraiment un sujet d'importance, je l'aurais probablement mis dans les candidats.</p>	<p>Critère de sélection: un sujet central du document.</p> <p>Critère de sélection pour un candidat descripteur: un terme essentiel du document qui ne peut être représenté par le vocabulaire du TH sera retenu en candidat.</p>
19	<p><u>La figure 2.1 présente leur emplacement respectif. Design des étangs. Étangs, je l'ai!</u></p> <p>E3: Oui. <i>Design des étangs</i>, on parle d'<i>étang</i>. Si une personne est intéressée à avoir des documents sur le <i>design des étangs</i>, elle va chercher avec <i>étang</i>. Ça me surprendrait qu'elle cherche avec <i>design</i>. De toute façon <i>design</i> n'est pas dans notre thésaurus.</p>	<p>Usager: Perception de la stratégie d'interrogation de l'usager.</p> <p>Utilisation de sa connaissance du thésaurus. Mention de l'impossibilité de représenter le terme <i>design</i> avec le vocabulaire du thésaurus.</p>
20	<p><u>Les sites seront déterminés parmi 5 différents emplacements potentiels situés dans la municipalité de Piedmont. (les raye sur sa feuille) St-Sauveur-des-Monts, St-Sauveur-Paroisse, je vais enlever ça.</u></p> <p>expé: Sais-tu ce qui t'a décidé? E3: <u>Emplacements potentiels situés (elle met l'accent sur ce dernier mot) dans la municipalité de Piedmont.</u> Donc on voyait que les <i>emplacements</i> étaient dans la municipalité de <i>Piedmont</i>, pas à <i>St-Sauveur-des-Monts</i>,</p>	<p>Confirmation du lieu géographique dont il est vraiment question dans le document.</p> <p>Ici, élimination des DEgéo non pertinents, c'est-à-dire qui ne concernent pas les emplacements.</p>

	<i>St-Sauveur-Paroisse.</i>	
21	<p><u>qu'aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs aérés. Station d'épuration, je vais l'enlever aussi (raye station d'épuration).</u></p> <p>expé: Pourquoi t'as enlevé station d'épuration? E3: Ici (<i>dans la conclusion</i>) on disait, <u>aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs.</u> Donc vraiment, le document porte sur l'<i>implantation d'étangs</i>, pas sur la <i>construction d'une usine</i>. C'est pour cela que je l'ai enlevé.</p>	Retenir ce dont le document traite vraiment. Ici, rejet des termes dont il n'est pas vraiment question.
22	<p><u>Contraintes d'évaluation des sites, contraintes environnementales. Grille d'évaluation, évaluation détaillée des sites. Évaluation des sites.</u></p> <p>expé: Ce n'est pas un terme qui est très général? E3: Oui, c'est un terme qui est très général. Dans le fond, on pourrait s'interroger sur la valeur d'avoir ce terme là dans notre thésaurus. Mais là, il est là, et vu que ça irait bien avec ce document-là, je l'ai mis. Puis ici (<i>dans la table des matières</i>) on voit la grosse partie qui parle de l'<i>évaluation des sites</i>.</p>	L'ampleur du traitement évoqué comme critère de sélection.
23	Si quelqu'un fait une recherche vraiment poussée et veut avoir l' <i>évaluation d'emplacement</i> pour telle construction, peut-être qu'il va le mettre. Si une personne cherche seulement avec <i>étang aéré</i> , elle va retrouver toute sorte d'affaires sur les <i>étangs aérés</i> : les répercussions environnementales, situés dans telle municipalité. Mais si elle trouve que c'est trop et elle veut spécifier, elle veut l' <i>évaluation</i> soit des <i>emplacements</i> ou des <i>étangs aérés</i> , elle peut spécifier avec.	Usager: Prise en compte d'une demande potentielle très spécifique.
24	Mais je ne suis pas sûre qu'une personne va chercher aussi spécifique que ça. Ça peut arriver, mais dans l'immédiat peut-être pas parce que peut-être qu'en cherchant juste avec <i>étang aéré</i> , il n'y aura pas tant de documents que ça qui vont sortir. Elle n'aura peut-être pas à fouiller tant que ça pour trouver ce qui l'intéresse.	Usager: Perception du type d'interrogation de l'usager. Référence au fonds documentaire actuel de la base Envirodoq.
25	<p><u>Bon! (<i>regarde sa feuille</i>) Mon document porte sur l'évaluation des emplacements... pour les étangs aérés.</u></p> <p>expé: Qu'est-ce que tu fais là? E3 Je fais ma petite phrase. Mon document, il parle de ça. Là, ça devait tirer à sa fin. En dernier, je fais une petite révision finale. Mon document, il parle de ça et je me dis aussi, si une personne vient et cherche là-dessus, il ne sera pas mécontent d'avoir ça. Ça va faire son affaire. En dernier, je reviens tout le temps là-dessus.</p>	Procédure d'indexation: construction d'un énoncé avec les descripteurs de sa liste et disposition de ses descripteurs dans cet ordre. Procédure d'indexation: révision finale et prise en compte de la satisfaction de l'usager.
26	<u>Traitement eaux usées, c'est plus général, je pourrais le mettre dans descripteurs secondaires.</u>	Définition des DE2: aspect plus général du document.

Concernant sa procédure d'indexation, on remarque que E3 procède en deux temps en choisissant d'abord tous les termes qu'elle juge pertinents lors de la prise de connaissance

du contenu et en éliminant par la suite certains de ces termes, en fonction de l'ensemble du contenu du document (extrait 2). Les extraits 20 et 21 témoignent d'ailleurs de l'application de cette procédure où E3 élimine les termes *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse* et *station épuration*. Par ailleurs, elle n'attribue pas de statut définitif à ses termes dans un premier temps (extrait 12). Lorsque sa connaissance du langage lui permet de traduire directement un terme en descripteur, elle effectue alors conjointement les étapes de sélection et de traduction (extrait 9). À la fin de son analyse (extrait 25), E3 ordonne ses descripteurs en un énoncé significatif et révisé sa liste en tenant compte alors de la satisfaction de l'utilisateur.

Cette indexation semble indiquer que E3 retient essentiellement l'information se rapportant au propos central du document (extraits 7, 8, 18, 20), laissant de côté les aspects plus secondaires (extraits 6, 16). Ceci dénote une indexation basée principalement sur les éléments textuels représentant la macrostructure du document (titre, titres principaux de la table des matières) (extrait 3), pouvant être associée à une procédure ne visant pas nécessairement l'exhaustivité. Le but du document, ou le propos central, semble être une dimension privilégiée par E3 (extraits 10, 14). Par ailleurs, l'ampleur du traitement et le nombre d'occurrences d'un terme sont évoqués fréquemment pour la sélection de termes (extraits 5, 7, 8, 15, 22), ce qui peut être associé à la prise en compte de l'exactitude.

E3 se sert de sa connaissance du thésaurus pour effectuer l'étape de traduction conjointement à celle d'extraction et semble privilégier les termes qui peuvent être représentés par le vocabulaire du thésaurus (extraits 1, 9, 15, 17, 19, 22). Il reste à déterminer si cette procédure est appliquée au détriment des critères de qualité ou si elle répond aux impératifs de la tâche et des contraintes du langage. Cependant, la représentation d'un terme jugé essentiel semble primer sur la sélection en fonction de sa présence dans le thésaurus, pour lequel un candidat sera alors inscrit (extrait 18). Quant à l'élément usager, il semble être pris en compte plus vers la fin du processus (extraits 23, 24) ou évoqué comme argument complémentaire pour renforcer une décision (extrait 11). On remarque également que E3 infère le type de demandes potentielles en rejetant le terme *design* (extrait 19) puisque selon elle, l'utilisateur ne questionnera pas la base avec ce terme. Soulignons que ce terme avait déjà été jugé non pertinent pour la représentation du contenu

(extrait 6), ce qui corrobore que la dimension usager est probablement évoquée en second pour l'évaluation d'un terme.

### 2.3 Prise en compte des éléments de connaissance par les indexeurs.

La section suivante présente successivement chaque élément de connaissance et en caractérise l'utilisation par chaque indexeur. Un tableau (Tableau XVI) synthèse est ensuite présenté afin de comparer les indexeurs et faire ressortir les différences interindexeurs majeures en rapport à certains éléments.

Tout d'abord, rappelons que la référence à un même élément de connaissance peut être exprimée sous différentes formes verbales par différents indexeurs ou différemment par un même indexeur en des moments différents. Ainsi, on ne doit pas s'attendre à reconnaître l'utilisation d'un élément de connaissance exclusivement par une mention explicite de cet élément tel que nommé dans les normes. Aussi, il arrive parfois que l'identification d'un élément de connaissance soit rendue possible grâce à l'examen du contexte entourant l'utilisation de cet élément (c'est-à-dire un épisode dans lequel s'insère l'énoncé) puisque l'énoncé seul ne permet pas de déterminer sans ambiguïté l'élément employé.

Soulignons que les énoncés cités en exemple peuvent comporter la mention d'autres éléments de connaissance, qui ne sont toutefois pas considérés à cette section. La première étape de la présentation des résultats vise à démontrer l'utilisation individuelle des éléments par chaque indexeur et ne tient pas compte de la combinaison d'éléments ou de la nature des liens entre les éléments de connaissance dans le processus global d'indexation. Ce dernier volet est couvert à une section ultérieure.

Notons finalement que dans le but de faciliter la lecture des énoncés, nous avons éliminé certaines particularités du discours parlé telles que redondances, suspension de phrases, etc., sans toutefois masquer ou déformer le sens de l'énoncé. Par ailleurs, le lecteur remarquera que la proportion des énoncés extraits des verbalisations consécutives (v2) est plus importante comparée à celle des verbalisations concomitantes (v1). En effet,

nous privilégions les énoncés qui illustrent de manière succincte et explicite le recours à un élément de connaissance donné ou à une stratégie particulière. Dans les v2, l'indexeur évoque l'utilisation d'un élément ou d'une stratégie en quelques phrases, ce qui permet d'illustrer le point sans alourdir le texte par la sélection de grandes portions de v1. Le recours aux v1 pour l'illustration des stratégies et appréhensions des éléments de connaissance nécessiterait des paragraphes entiers afin de situer la décision (choix, rejet ou évaluation d'un terme) de l'indexeur dans son contexte global afin que le lecteur saisisse le lien entre ce que l'indexeur verbalise et l'argumentation avancée. Ainsi, nous présentons des v1 seules, lorsque le rapport avec l'élément de connaissance est clair; des v1 accompagnées de v2; ainsi que des v2 seules, toujours avec le souci de faciliter la compréhension du lecteur dans cette complémentarité argumentation et illustration.

### 2.3.1 Exhaustivité

E1

Les listes brouillon de E1 dénotent une représentation élaborée du contenu du document qui est traduite en une liste de descripteurs considérablement plus concise. Les arguments avancés par E1 font ressortir que l'exhaustivité ne semble pas prévaloir pour la sélection des descripteurs.

#### **E1 Étang aéré v2**

C'est mieux d'en mettre moins que de me perdre et en mettre trop, que ce ne soit pas valable.

#### **E1 Biogaz v2**

Comme là, il y en a deux (*elle extrait les termes caractérisation et élimination pour représenter les actions sur le biogaz*). Si on prend *caractérisation*, ça englobe *composition* et *générations*. *Élimination* c'en est un autre. Deux, ce serait suffisant.

#### **E1 Biogaz v2**

Si t'en as trop ça devient incohérent un moment donné. T'es aussi bien de te fier au titre de ton document et ça va être beau pour le repérer. Il faut se mettre à la place du chercheur (*usager*). Si on fait cela pour chaque document qui existe dans la banque de données, il va se retrouver avec une recherche, admettons qu'il demande élimination déchet, il va y sortir deux cent documents.

D'ailleurs, E1 souligne en entrevue qu'elle favorise l'exactitude plutôt que l'exhaustivité afin d'éviter de submerger l'utilisateur avec des notices non pertinentes; ce qui justifie en partie la concision de ses listes de descripteurs.

De plus, la non-exhaustivité de ses listes peut également être justifiée par les références fréquentes à des inclusions, comme l'indiquent les énoncés suivants.

**E1 Étang aéré v1**

Pour le moment, *épuration* n'est pas nécessaire, c'est compris dans *traitement des eaux usées*.

**E1 Biogaz v2**

C'est un terme important (*incinérateur*), mais il est beaucoup lié à *élimination déchet*. Il me semble que *incinération déchet*, ça fait partie de *l'élimination de déchet*.

**E2**

E2 tient à représenter de manière exhaustive le contenu du document. Pour ce faire, elle cherche à attribuer un descripteur pour chacune des facettes (aspects) abordées, en respectant une limite prédéterminée de six descripteurs pour chacun des champs DE1 et DE2.

**E2 Étang aéré v2**

À la fin, je regarde ce que j'ai mis et passé six (*pour chaque niveau*), j'essaie d'en enlever. C'est rare que je dépasse ça.

**E2 Biogaz v2**

Avec mes DE1, je pense avoir fait ressortir les sujets primordiaux. Les petits à-côtés, c'est les DE2 que je trouve quand même importants qui sont mentionnés dans le document.

Son application de l'exhaustivité semble toutefois être conditionnelle à l'ampleur du traitement du terme évalué. Il est possible que pour E2, le respect de l'exhaustivité soit atteint lorsque tous les aspects importants, ou traités substantiellement, sont représentés.

**E2 Étang aéré v2**

À la fin, je regarde de nouveau ce que j'ai mis sur la feuille et si ç'a vraiment rapport. Je regarde mes termes pour savoir si ça vaut vraiment la peine de les avoir mis. Pour moi, cerner le sujet, entourer le sujet, c'est comme une révision, si ça de l'allure. Et je me dis, ai-je vraiment tout sorti ce qu'il y avait d'essentiel par les termes?

**E2 Énergie v2**

**expé**

*Climat* tu viens de le rayer (*de sa liste brouillon*).

**E2**

Oui. Il n'en parle pas assez.

## E3

Peu d'énoncés de E3 se rapportent explicitement à l'exhaustivité. Cependant, les listes de descripteurs de E3 témoignent d'une prise en compte moindre pour l'exhaustivité. En fait, E3 ne retient que les termes qui se rapportent au sujet central et par conséquent écarte généralement les aspects secondaires de sa représentation. Certains des termes jugés plus secondaires sont considérés inclus dans le terme principal auquel ils se rapportent.

**E3 Compostage v2**

Je les (*descripteurs*) pondère différemment parce que je trouve certains plus importants que d'autres. Mais je les mets tous en DE1, parce que je trouve que le document c'est tout ça. Tandis que DE2, c'est plus par partie.

Notons cependant que ses listes de descripteurs contiennent peu de DE2, tout comme les listes de E1.

**E3 Énergie v2**

Je me demandais si c'était pertinent (*éclairage*). Mais non. Parce que je n'ai pas *cuisson*, je n'ai pas *puits*. Donc je ne mettrai pas *éclairage* plus que les autres. Je trouvais que c'était trop dans le détail.

Ce dernier énoncé fait ressortir le niveau d'indexation recherché par E3 et souligne l'opposition entre E3 et E4 pour l'utilisation de l'exhaustivité. Alors que pour E3 les termes *éclairage*, *cuisson* et *puits* sont rejetés parce que trop spécifiques pour la représentation du document, E4 les juge intéressants et les intègre à sa représentation.

**E4 Énergie v2**

*Cuisson* et *éclairage*, on fait référence à ça assez longuement dans la table des matières. Et là je me mets dans le contexte que c'est un village de quatre cent quatre-vingt-seize habitants au Sénégal, qui effectivement, cet aspect-là est important pour eux.

## E4

Le nombre de descripteurs contenus dans les listes de E4 démontre sans contredit un souci d'exhaustivité, devant toutefois être défini. L'application de l'exhaustivité chez E4 est associée à un relevé exhaustif de toutes les notions et les sous-aspects de ces notions présentés dans le document sans que la pertinence de ces termes soit remise en question. Précisons que parmi les critères de qualité utilisés par E4, l'exhaustivité est privilégiée.

**E4 Étang aéré v2**

C'est évidemment les mots en caractères gras qui séparent les petits chapitres. Ils ont été écrits en caractères gras c'est parce qu'ils ont une certaine importance dans le document. Alors je vais m'arrêter à ces mots-là plus.

**E4 Étang aéré v2**

Ça ne me viendrait jamais à l'idée de mettre dix descripteurs principaux.

**expé**

Mais plus en secondaires?

**E4**

Oui. Parce que si tu veux aller chercher un peu tous les aspects dont il est question, en rapport avec les descripteurs principaux, ça peut t'amener à en sortir beaucoup de termes.

En entrevue, E4 souligne effectivement qu'elle accorde la priorité à cet élément afin de n'omettre aucune notion qui passerait alors sous silence au repérage.

Il est important de souligner que E4 mentionne certains principes à suivre pour l'indexation, mais ne les applique pas concrètement. Il semble donc y avoir contradiction entre les verbalisations et les actions de E4. Cependant, puisque l'on a été en mesure de repérer les différences entre les principes verbalisés et les comportements réels, nous pouvons néanmoins identifier les éléments de connaissance utilisés pour le choix de termes.

#### Contradiction verbalisations-actions de E4

Premièrement, E4 argumente que son relevé exhaustif de termes à l'étape d'extraction consiste à se donner plusieurs pistes et options pour le choix des descripteurs finaux.

**E4 Étang aéré v2**

Je me donne du matériel pour faire une espèce d'épuration après, des termes que j'ai inscrits. Pour essayer de garder vraiment ceux qui sont plus importants que d'autres.

**E4 Énergie v2**

Je regarde ma feuille et j'en mets des mots hein! Pour avoir la possibilité d'aller en trouver le plus possible après dans le thésaurus. Je m'assure que j'en ai mis assez pour être capable de les rendre en descripteurs.

**E4 Biogaz v2**

On a parlé d'*émission*, d'*odeurs*, de *retombées atmosphériques*. *Qualité air*, c'est une autre piste pour aller chercher la même information.

Cependant, on observe que pour le document *Biogaz*, E4 conserve *odeur* et *qualité air* en descripteurs et élimine *émission* et *retombées atmosphériques* pour lesquels elle n'a pas trouvé d'équivalent dans le thésaurus.

Afin d'illustrer et de caractériser son souci permanent d'exhaustivité aussi bien à l'étape d'extraction que de traduction, nous présentons une partie du protocole qui se rapporte à l'extraction de termes à partir de la lecture du résumé<sup>6</sup> du document *Énergie*<sup>7</sup>.

#### E4 Énergie v1

Un résumé: Le village de Ndia Gouréyé est situé au nord du Sénégal, dans la zone climatique sahélienne. Peut-être Sahel. Le climat est caractérisé par une très forte variation des pluies et par des précipitations annuelles comprises entre 400 et 600 mm de hauteur d'eau. La période de l'hivernage y dure environ trois mois. On peut parler de climat, la pluie. La période de l'hivernage. Saison, peut-être? Pendant la saison sèche et notamment aux mois de décembre et janvier, un vent froid et sec, le harmattan, y souffle du N-E avec beaucoup de vigueur. Vent. Le sol est surtout constitué d'éléments sablonneux. Sol. Les sols dior, sur lesquels poussent facilement l'arachide. L'ensoleillement du site y est très important, ce qui accroît le phénomène de désertification... En tout cas, ensoleillement, désertification.... très important dans cette région sahélienne. La végétation naturelle de la région est constituée d'un réseau clairsemé d'acacia, de soump, de Kad, de baobab et de tamarinier qui progressivement depuis de nombreuses années, a laissé place aux champs d'arachides, de mil, de sorgho et de manioc. Alors végétation. OK. Les structures agricoles... Alors, je mets agriculture. ...de la région ont été influencées par la rigueur du climat et la pauvreté des sols, mais aussi par les conséquences qu'a entraîné la colonisation, notamment en implantant la monoculture arachidière au détriment de la polyculture traditionnelle... Monoculture. ...polyculture traditionnelle de subsistance et en modifiant le mode d'appropriation du sol qui était plutôt collectif pendant la période pré-coloniale. Aujourd'hui, dans ce village de 496 habitants vivant pauvrement d'agriculture et d'élevage de subsistance, les problèmes liés à l'exhaure de l'eau, Ça, faut que j'écrive ça: exhaure de l'eau. ...à la cuisson des aliments et à l'éclairage sont cruciaux. Cuisson, aliments, éclairage. Les solutions proposées afin de résoudre ces problèmes sont les suivantes: utilisation d'une éolienne servant à l'exhaure de l'eau. Alors éolienne. L'implantation d'un système de cuisinière solaire pour compenser la rareté du bois utilisé actuellement pour la cuisson des aliments qui est une activité très pénible. Bon! Énergie solaire. Recyclage de la fumure animale en biogaz devant servir à produire l'électricité nécessaire à l'éclairage. Un recyclage, déchet animal, biogaz. ...devant servir à produire l'électricité nécessaire à l'éclairage. Je mets électricité avec production entre parenthèses. Cette recherche se veut être avant tout une étude de faisabilité afin de démontrer qu'il est possible et viable d'utiliser les énergies renouvelables,...

Y avait l'eau, problèmes de l'eau, les puits....

Une présentation de l'équipement proposé ainsi qu'un scénario possible suivent ces deux premières parties pour résoudre les problèmes énergétiques. Bon! Déjà, j'ai l'impression d'avoir tous les mots dont j'ai besoin. Il s'agit de les retrouver dans le thésaurus.

À l'étape de traduction, E4 reprend un à un les termes inscrits et cherche le descripteur approprié dans le thésaurus. Elle n'évalue pas les termes en fonction de leur représentativité du contenu du document avant d'en faire la traduction. En fait, E4 ne procède pas au "débroussaillage" mentionné en verbalisations consécutives, mais transpose

<sup>6</sup> Afin de réduire la longueur de cet épisode, nous avons retiré quelques paragraphes pour lesquels E4 ne retient aucun terme.

<sup>7</sup> Rappelons que les caractères soulignés se rapportent à l'activité de lecture, les caractères gras à l'activité d'écriture et le texte normal à une activité de traitement verbalisée.

systématiquement tous les termes pour lesquels elle trouve un équivalent dans le thésaurus, et ce pour l'indexation des quatre documents.

Voici quelques exemples de traduction des notions retenues à la première étape pour l'indexation du document *Énergie*:

<b>Termes extraits</b>	<b>Descripteurs retenus</b>
Climat	climatologie, changement climatique
Pluie	précipitations
Vent	vent
Désertification	désertification, zones arides, sécheresse
Électricité (production)	énergie électrique
Saison, cuisson, aliments	<i>non traduits</i>

Il arrive que plus d'un descripteur soit jugé valable pour représenter une même notion. Dans ce cas, E4 ne cherche pas à retenir le descripteur le plus approprié et conserve alors deux ou trois descripteurs associés. Par exemple, les descripteurs *habitation*, *zonage* et *expropriation* (mis en candidat descripteur) sont reliés au concept des *contraintes environnementales* pour le document *Étang aéré*; les descripteurs *législation* et *règlement* représentent la notion de lois pour le document *Biogaz*; et pour le document *Compostage*, les descripteurs *technique de conversion des déchets*, *recyclage* et *récupération des déchets* sont associés au concept de la valorisation des boues (eaux usées).

Cette multiplication des entrées sur la liste finale témoigne une fois de plus de la prédominance de l'exhaustivité tout au long du processus d'indexation de cet indexeur. À titre comparatif, soulignons que E1 représente la notion *d'énergies renouvelables* du document *Énergie* par le DE *ressources renouvelables*, E2 utilise le DE *utilisation des ressources locales* et E3 choisit le DE *source d'énergie non-polluante*, tandis que E4 retient ces trois descripteurs.

Par ailleurs, les termes pour lesquels E4 ne trouve pas d'équivalents dans le thésaurus sont généralement abandonnés, sans autre considération, sauf s'il s'agit d'un terme inusité ou d'actualité.

Cette application très particulière de l'exhaustivité par E4 semble s'effectuer au détriment de l'exactitude et de la précision.

En somme, ceci nous permet de constater que les deux indexeurs les plus exhaustifs ont une appréhension différente de ce critère de qualité. Alors que E2 se limite à l'utilisation d'un descripteur pour chaque aspect ou facette du document en misant sur l'ampleur du traitement (liée à l'exactitude), E4 retient la plupart des termes présents dans les parties les plus informatives et mis en évidence par une typographie particulière dans le corps du texte, sans évaluer la valeur réelle de ces termes pour la représentation du contenu du document.

Ceci explique d'une part, que E2 et E4 sont les seuls à retenir certains termes, tel que *coût* pour le document *Étang aéré*, *échantillonnage* pour le document *Biogaz* et *éclairage* pour le document *Énergie*, et d'autre part, que E2 rejette certaines notions (*bruit*, *méthane*, *règlement*, *législation*) conservées par E4.

### 2.3.2 Précision et Exactitude

Ces deux critères étant étroitement liés autant dans la pratique que dans la description qu'en donne la norme ISO (1985), nous les considérons conjointement dans la procédure d'indexation de chaque indexeur. Pour pouvoir caractériser une prise en compte de la précision ou de l'exactitude, nous devons souligner les particularités qui permettent de distinguer ces deux éléments.

Ainsi, nous associerons l'application de la précision avec le niveau d'indexation recherché par l'indexeur, c'est-à-dire le fait qu'il vise à indexer au même niveau de traitement que celui présenté dans le document. Toutefois, une attention particulière sera portée aux contraintes imposées par les vocabulaires utilisés qui ne permettraient pas l'application de cet élément. Il sera donc judicieux de distinguer le moment de la prise en compte, soit l'étape d'extraction ou l'étape de traduction, puisqu'un indexeur pourra tenir compte de la précision, mais ne pourra peut-être pas la respecter lorsque confronté aux possibilités offertes par le langage.

L'exactitude sera considérée comme prise en compte lorsque l'indexeur vise à retenir des notions qui sont effectivement présentes dans le document, c'est-à-dire que le document en traite suffisamment pour justifier la sélection du terme ou que l'indexeur cherche à représenter une notion ou un terme par un descripteur qu'il sait être adéquat; ce qui fait référence à sa connaissance du domaine.

## E1

Concernant la précision, E1 ne semble pas s'attarder à la correspondance du niveau hiérarchique entre termes et descripteurs. En fait, E1 retient plus un DE sur la base de l'adéquation du sens du terme avec celui d'un DE. Les rares mentions de la précision faites par E1 à l'étape de traduction soulignent que le descripteur n'est pas aussi précis que le terme original. Toutefois, cette dimension n'influence pas sa décision, c'est-à-dire que lorsque E1 juge qu'une notion doit être représentée, son choix ne repose pas sur la précision du descripteur, mais plus sur son exactitude. Par ailleurs, on remarque que d'autres éléments de connaissance sont évoqués conjointement avec la précision et semblent avoir une influence marquée sur la décision finale.

Par exemple, les énoncés suivants soulignent que l'évolution du domaine et l'utilisateur semblent être des facteurs importants pour le choix de descripteurs.

### E1 Biogaz v1

(*Regarde sa liste*) Torchère à flamme visible, ici. On va juste mettre *torchère à flamme*, tout seul... (raye invisible)

#### expé

Tu décides de garder que *torchère à flamme* au lieu de garder les deux. Pourquoi?

### E1 Biogaz v2

La personne qui cherche *torchère à flamme*, si elle connaît juste les *torchères à flammes invisibles*, elle va probablement chercher de cette façon là. Mais il va sortir pareil *torchère à flamme*, tu mets des astérisques. Comme il est tellement spécifique. Au fur et à mesure, au bout d'un an je vais en avoir indexé plusieurs documents avec *torchères à flammes visibles* tout seul. Alors à ce moment je change mon indexation; là ce sera peut-être important de le préciser (*sur sa liste finale, E1 conserve torchère en candidat DE*).

### E1 Biogaz v2

(**Informatique**) C'est quand même général. Mais par contre, par expérience, je sais qu'aussitôt qu'il existe un logiciel applicable à l'étude ... c'est important de le mettre parce qu'il y en a qui vont faire de plus en plus juste leur recherche sur les logiciels. Donc juste en mettant *informatique*, ils (*usagers*) vont savoir que c'est sûrement un logiciel ou une application informatique de quelque chose.

En fait, il semble que la précision joue un rôle négligeable dans ses décisions. Ainsi, puisqu'il apparaît que la précision peut être mentionnée, mais demeure non décisive pour le

choix de termes, il nous semble approprié de placer ce critère parmi les éléments peu ou pas utilisés par E1.

L'application de l'exactitude par E1 comporte deux volets. Premièrement, l'exactitude est évaluée en fonction du nombre d'occurrences ou de l'ampleur du traitement des termes traités qui se retrouvent en majorité dans les parties les plus informatives du document.

#### **E1 Biogaz v2**

(*À propos des termes combustion et destruction rattachés au sujet principal biogaz*) Si ça revient souvent je vais faire plus attention, je vais essayer de le mentionner.

#### **E1 Énergie v2**

Ça (*énergies renouvelables*) se trouve à être un synonyme des différentes énergies qui sont mentionnées dans le titre. Et tu le trouves ici aussi (*dans le résumé*). (*Elle l'inscrit en brouillon*.)

Parallèlement, E1 s'assure que le document traite réellement des termes qu'elle veut retenir et en conséquence, rejette les notions qui sont mentionnées mais non développées. En plus des termes soulignés dans les énoncés suivants, cet argument rend compte du rejet des termes *incinération* et *effet serre* pour le document *Biogaz*, lesquels, par comparaison, sont retenus par E4.

#### **E1 Étang aéré v1**

Alors, finalement ça m'a l'air d'être une étude préliminaire. C'est une étude qui est présentée au ministère. Il n'y a pas de décision de prise, y a pas d'analyse de coûts (*rejette la notion de coût*).

#### **E1 Étang aéré v2**

(*À propos du terme dézonage*) Mais s'ils en avaient parlé plus longuement. Il aurait pu y avoir une suite à l'étude. Une annexe qui aurait pu être faite sur les effets du *dézonage*. Mais ils font juste le mentionner de même. Je l'ai marqué, je l'ai laissé tomber ... c'est trop loin du concept qu'on a déjà sorti, *traitement eau usée*. C'était plutôt encore en relation avec *usine de biofiltration* que j'allais voir *dézonage*. Mais *dézonage*, ça peut être pour n'importe quoi. Non.

Deuxièmement, E1 semble porter une attention particulière au sens véhiculé par les termes du document et le sens des descripteurs susceptibles de les représenter. D'ailleurs, E1 souligne en entrevue que lorsqu'elle n'est pas certaine de l'adéquation du sens du terme avec celui du descripteur, elle préfère rejeter le descripteur plutôt que de créer une fausse piste pour le repérage.

#### **E1 Étang aéré v2**

Je n'ai pas le temps de voir qu'est-ce que ça veut dire exactement *épuration biologique*. Peut-être que ce serait un bon terme. Peut-être que c'est exactement de ça dont on parle. Mais il me semble qu'il serait apparu dans le document, si ç'avait été exactement ça.... Mais comme je ne suis pas sûre, je laisse faire.

**E1 Compostage v2**

*(E1 rejette le descripteur récupération des déchets pour représenter le concept de la valorisation des boues, la réutilisation des eaux usées) Je sais qu'en réécoutant ce que je disais, déchets récupération, c'est (le document porte sur une) une forme de récupération des déchets. Mais quand on pense à l'expression récupération des déchets, on voit ça comme les cannes et de la vitre, c'est autre chose.*

**E1 Compostage v2**

*C'est un terme bien spécifique, épandage, s'il n'est pas là, il n'est pas là (si le thésaurus ne contient pas ce terme exact, aucun autre ne peut le remplacer).*

Finalement, soulignons qu'à travers la description de son appréhension de la précision et de l'exactitude, la prise en compte de l'usager revient régulièrement, dimension qui sera élaborée à la section correspondante.

**E2**

E2 tient à préciser l'angle sous lequel est abordé le ou les thèmes principaux en représentant chacun des aspects du document, ce qui parfois la conduit à enfreindre le respect de la précision telle que définie par la norme. Ainsi, lorsque seul un descripteur plus générique permet de rendre compte d'un aspect jugé important, la spécificité sera abandonnée. En fait, il serait plus juste de dire que E2 ne mise pas en priorité sur une représentation précise et accepte de faire un compromis de précision afin de représenter le terme jugé pertinent pour la représentativité du contenu du document.

**E2 Biogaz v2**

*(biogaz n'étant pas dans le thésaurus, E2 cherche un DE pour le représenter)*

*Je garde gaz carbonique quand même. Parce que pour moi biogaz, c'était des gaz carboniques. Une forme de gaz carbonique, mais sous une forme générale, pas spécifique.*

**E2 Étang aéré v2**

*De ne pas trouver ce que je veux c'est frustrant. De proposer autre chose, ce n'est pas frustrant du tout. (Lorsque le terme n'a pas son équivalent exact dans le thésaurus) mais là je m'arrange pour trouver autre chose, un autre terme, des synonymes. Je vais faire plus de démarches c'est sûr.*

Cependant, E2 tient à une représentation exacte du contenu du document, en considérant un terme exact lorsque le document en traite suffisamment et que le sens n'est pas trahi par un descripteur inapproprié. Ainsi, afin de s'assurer de l'adéquation du sens des deux mots (terme et descripteur), E2 vérifiera le contexte dans lequel le terme du document est utilisé et celui du descripteur en relation avec les termes qui l'entourent (TS, TG, TA).

Soulignons que la satisfaction de l'utilisateur est parfois évoquée conjointement à l'évaluation du sens du terme et du descripteur pour décider du choix ou du rejet d'un descripteur.

### **E2 Énergie v1**

Plus on s'avance à l'intérieur des terres et plus la nappe phréatique est profondément enfouie. Aussi les puits ont souvent une quarantaine de mètres de profondeur. Puiser de l'eau. Oui! **Approvisionnement en eau dans les zones rurales**, ça je garderais ça.

### **E2 Énergie v2**

Je cherchais juste à vérifier si c'était vraiment le bon terme.

### **E2 Biogaz v2**

*(Caractérisation n'étant pas dans le thésaurus, E2 évalue l'équivalence du descripteur identification)*

*Identification*, c'est plus le nom qu'on donne au fait de trouver quel genre de gaz, identifier. *Caractérisation* c'est plus que cela, c'est comment se comporte ce gaz, c'est plus le caractère propre au gaz. Donc, ce n'est pas la même chose du tout. Ça ne me satisfait pas. Mais je savais que si je mettais ça, je ne serais pas contente du tout. Quand ça ne me satisfait pas, j'aime quasiment mieux rien mettre que mettre un terme qui est mauvais, qui ne veut pas dire la même chose.

Cette application de l'exactitude jointe à une considération moindre pour la précision justifie le choix ou le rejet de plusieurs descripteurs. Ainsi, *techniques de conversion des déchets et développement économique* pour le document *Compostage et approvisionnement en eau dans les zones rurales* pour le document *Énergie* ont été sélectionnés sur la base de la correspondance du sens avec le terme du document, malgré leur manque de précision. De même, le descripteur *usine de traitement des eaux d'égouts* est rejeté parce qu'il ne rend pas le sens du terme du document.

Aussi, pour représenter le plus exactement possible le contenu du document, E2 pourra ajouter des descripteurs pour spécifier le sujet principal. Entre autres, ceci permet d'expliquer le choix des descripteurs *station épuration, traitement eau usée et municipalité* pour le document *Étang aéré*. Notons que cette appréhension plus globale de l'exactitude s'apparente à l'exhaustivité.

## **E3**

Premièrement, E3 cherche à représenter le contenu du document de manière la plus exacte possible en privilégiant les termes dont le document traite amplement et en vérifiant l'adéquation du sens du terme avec celui du descripteur potentiel.

Les deux énoncés suivants soulignent l'importance de l'ampleur du traitement pour le choix d'un terme. *Évaluation* est conservé en partie parce que le document en traite amplement et *équipement* n'est pas retenu puisque ce n'est pas une notion développée dans le document.

### **E3 Étang aéré v2**

*Évaluation*, c'est un terme qui est très général. Dans le fond, on pourrait s'interroger sur la valeur d'avoir ce terme-là dans notre thésaurus. Mais là, il est là, et vu que ça irait bien avec ce document-là, je l'ai mis.... Et ici, on voit la grosse partie qui parle de *l'évaluation des sites*.

### **E3 Énergie v2**

Peut-être *équipement*, la présentation de l'équipement. S'il y avait une bonne partie qui traitait de *l'équipement*, j'aurais peut-être développé là-dessus. (E3 l'a écrit en brouillon et ne le retiendra pas puisqu'il n'est pas développé).

L'application de l'exactitude doit également répondre à l'adéquation du sens entre terme du document et descripteur, tel que le démontre l'énoncé suivant.

### **E3 Étang aéré v2**

Dans ma tête ça s'est passé, *contraintes* (le terme du document est *contraintes environnementales*) et *répercussions* (le DE *répercussions environnementales*), ce n'est pas la même chose. On parle des *contraintes*, mais pas des *répercussions* sur l'environnement de faire des sites. Si on en parle c'est une phrase par-ci par-là, mais je ne suis pas allé voir dans le détail parce qu'il n'y avait pas de grand thème qui parlait de *répercussions*. Alors tout de suite je l'ai abandonné.

Cette appréhension de l'exactitude explique également le rejet des DE *station épuration*, *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse* pour sa représentation du document *Étang aéré*, lesquels ont été retenus par les autres indexeurs. En effet, E3 rejette ces notions car le document ne porte pas spécifiquement sur ces aspects.

De plus, une représentation précise des notions traitées est également visée par la sélection du descripteur qui s'apparente le plus au niveau du terme à représenter.

### **E3 Biogaz v2**

J'essayais de débroussailler le terrain parce que j'avais mis toute sorte d'affaire. De quoi ça parle au juste et précisément.

### **E3 Énergie v2**

(E3 choisit entre les descripteurs *énergie de la biomasse* et *bois comme combustible*)

expé

Même si biomasse était dans le titre, tu préférerais bois comme combustible?

**E3**

Oui, c'était plus précis. Mais après, je suis retournée dans le document et c'est vraiment des bouses, humaines ou animales. C'était plus précis que de la *biomasse*. *Biomasse* c'est plus vague, c'est (dans le thésaurus) déchet employé comme combustible, terme spécifique (*TS au DE biomasse*).

L'épisode suivant illustre le processus de réflexion qui accompagne l'application de la précision chez E3. Ce sont les notions *élimination des déchets* et *enfouissement sanitaire* qui sont évaluées.

### **E3 Biogaz v1**

Tecsult a été mandaté par la Ville de Montréal pour sélectionner une technologie adéquate de destruction du biogaz produit par le site d'enfouissement du Centre de tri et d'élimination des déchets. *Élimination des déchets* là. (*raye élimination des déchets sur sa feuille brouillon*)

### **E3 Biogaz v2**

Le document porte sur le *biogaz* qui est produit par *l'enfouissement des déchets*. *Élimination des déchets*, ce n'est pas vraiment le sujet du document, c'est plutôt la cause de ce document. Quand on élimine des déchets dans un *site d'enfouissement*, on a ce problème-là, ce qui nous amène à faire ce document.

*Enfouissement sanitaire*, j'ai pensé que c'était plus approprié parce que s'il y a des *biogaz*, c'est parce qu'on fait de *l'enfouissement sanitaire*. Parce qu'on *élimine des déchets*, c'est plus lointain. La cause plus directe c'est que les *biogaz* sont produits par *l'enfouissement sanitaire*. C'est plus précis que *l'élimination des déchets*.

Ainsi, son niveau d'indexation recherché s'apparente à son évaluation de l'exactitude, les descripteurs retenus permettant un repérage de documents très pertinents puisque non seulement le document traite de ces notions précises, mais il en traite abondamment. En conséquence, il semble que E3 privilégie presque exclusivement une représentation du propos central du document.

Par ailleurs, E3 applique l'élément de précision pour distinguer les descripteurs principaux des secondaires.

### **E3 Biogaz v2**

J'essaie d'aller au plus précis. Quand c'est plus général, j'essaie de le mettre en descripteur secondaire.

### **E3 Étang aéré v2**

DE2 c'est plus les aspects par chapitre. Moi, je suis portée à mettre en DE2 des termes plus généraux. Ce n'est pas le sujet précis, en DE2. Mais on n'a pas vraiment de politiques écrites là-dessus.

Finalement, une façon indirecte d'identifier la prise en compte de ces critères de qualité consiste à repérer les mentions de compromis effectués en rapport à l'impossibilité de représenter les notions retenues de manière précise et exacte. Ainsi, l'insatisfaction explicite de E3 pour sa liste de descripteurs du document *Compostage* permet de souligner l'importance qu'elle accorde à ces critères. Ceci justifie sa décision de retenir *biogaz* et *compostage* en candidats descripteurs, puisque chacun de ces termes représente la notion primordiale du document.

On remarque dans l'exemple suivant que E3 cherche un descripteur dont le sens s'apparente le plus au terme du document à représenter. Ainsi, malgré les compromis inévitables, E3 cherche le descripteur le plus exact parmi le vocabulaire d'Infoterra.

#### **E3 Compostage v2**

Ici j'avais sorti les technologies de compostage. Ça pourrait parce que ça réduit les boues. *Compostage* pouvait être une *technologie de lutte contre la pollution*. Je trouvais que *moyens de* et *technologie de lutte*, ça se ressemblait.

#### **E4**

À l'étape d'extraction, l'application de l'exactitude chez E4 semble se limiter à transcrire les termes du document selon leur forme exacte. Le premier énoncé se rapportant au document *Étang aéré* illustre ce fait. E4 retient le terme *eaux usées*, tel qu'identifié à partir de la structure de surface du document, plutôt que l'expression *traitement eaux usées*, retenue par les trois autres indexeurs et figurant telle quelle dans le thésaurus:

#### **E4 Étang aéré v1**

*(Lecture de l'introduction) Les municipalités de Piedmont, St-Sauveur-des-Monts et St-Sauveur-Paroisse se sont regroupées dans le but d'effectuer un traitement commun de leurs eaux usées se déversant dans la Rivière du Nord. Eaux usées (inscrit sur feuille de travail).*

Son utilisation de la précision consiste à ajouter un terme qui spécifie le sens d'une notion principale déjà retenue, sans chercher à ne retenir que le terme le plus approprié. En fait, E4 semble entourer la notion de termes appartenant au même réseau sémantique, sans remettre en question la valeur de la précision qu'ils apportent à la représentation du contenu réellement traité dans le document. La seule présence d'un terme dans les parties les plus informatives du document apparaît suffisante pour que le terme soit retenu, du moins à la première étape de l'indexation.

#### **E4 Étang aéré v2**

*(Relié à la lecture de la conclusion) On avait parlé d'habitation, de zonage (deux termes qu'elle a retenus) et là on parle d'expropriation. C'est comme plus défini, plus précis.*

#### **E4 Compostage v2**

*(Écrit humus et amélioration sol qu'elle relie à valorisation agricole et compostage déjà inscrits) c'est en référence avec la valorisation au point de vue agricole. Ça va enrichir l'humus, ça améliore le sol. C'est d'autres termes.... Compostage, c'est pour améliorer la terre, donner un bon humus. Ce n'est pas dans le cœur (du document), mais ça ne serait pas loin.*

**E4 Énergie v2**

(*Écrit ensoleillement et désertification*) C'est juste par rapport au caractère physique de ce coin de pays là qui est situé au Sahel. Ce sont tous des phénomènes qui sont importants là-bas.

Cette appréhension de la précision et de l'exactitude ne correspond que très peu à la description de la norme. En effet, les termes ne semblent pas choisis en fonction de leur capacité à discriminer les documents les plus pertinents de l'ensemble des documents du fonds.

Si l'on peut mettre en doute une prise en compte réelle des critères de précision et d'exactitude à l'étape d'extraction, il semble clair qu'à l'étape de traduction, le choix de termes ne répond qu'à l'impératif d'exhaustivité. Une fois sa liste de termes complétée, E4 entreprend la transposition systématique de chacun de ces termes en descripteur sans questionner la spécificité et l'exactitude du descripteur choisi. Il s'ensuit que plusieurs termes sont représentés par deux descripteurs ou qu'un descripteur composé n'est retenu qu'en fonction d'un segment de son expression.

**E4 Étang aéré v2**

(*L'expression du document est contamination des puits*<sup>8</sup>) **Contamination et puits.** Si je mets puits tout seul, ça ne veut pas dire grand chose. Ce sont deux qui sont carrément rattachés l'un à l'autre. À cause du thésaurus, t'es obligé de faire cela des fois. T'es obligé d'aller chercher les deux mots, de les associer pour trouver ce que tu veux comme information quand tu fais ta recherche.

**E4 Compostage v1**

(*Consultation du thésaurus*) Non, je ne mettrais pas *système (d'épuration des eaux usées)*. Je vais mettre **usine de traitement des eaux d'égout...**

**E4 Compostage v2**

Moi, j'avais *station (sur sa feuille brouillon)*. *Usine* est beaucoup plus proche de *station d'épuration* que de *système d'épuration*. Je trouvais que c'était mieux de garder ça que l'autre, même avec les *eaux d'égout*. C'est vraiment pour le mot *usine* que je l'ai gardé lui.

Dans son processus global d'indexation, E4 n'évalue pas un terme en fonction de l'ampleur du traitement qui en est fait et elle ne vérifie pas l'adéquation du sens de la notion du document à celui du descripteur du thésaurus. Toutefois, la préoccupation de l'exactitude semble poindre, du moins implicitement, lors de l'attribution du statut des descripteurs. En effet, E4 place en DE1 les descripteurs qu'elle qualifie de "cœur du

<sup>8</sup> Soulignons qu'une seule phrase de l'ouvrage se rapporte à la contamination des puits.

document", lesquels correspondent aux termes dont le document traite amplement et sont issus principalement du titre et des grands-titres de la table des matières.

Cependant, bien qu'elle distingue les termes les plus importants en leur attribuant le statut de DE1, E4 ne compare pas la représentativité des notions afin de conserver les plus importantes et n'évalue pas la capacité de discrimination des termes. Par conséquent, ceci semble indiquer que l'exactitude et la précision ne sont pas réellement prises en compte.

En résumé, les notions principales (DE1) choisies par E4 semblent répondre aux critères de précision ou d'exactitude, mais sa préoccupation marquée pour l'exhaustivité laisse deviner une influence négligeable qu'exercent réellement ces éléments sur l'ensemble des choix de termes et descripteurs.

### 2.3.3 Usager<sup>9</sup>

#### E1

E1 tient compte de l'utilisateur tout au long de son indexation et affirme indexer autant pour l'utilisateur spécialiste que pour "monsieur-tout-le-monde".

#### **E1 Étang aéré v2**

Je vois autant un ingénieur qui peut faire une étude spécialisée sur des sites semblables ou pour le même genre de problèmes environnementaux. Et je vois autant aussi monsieur dans sa cour qui a un problème justement, on pense à ceux qui restent sur le bord des rivières ou des lacs. S'il ne connaît pas le titre, ou les termes techniques, il va chercher à Piedmont naturellement. Il faut mettre Piedmont et faut mettre le nom du site, ... peut-être que pour lui c'est son seul repère (*les noms de sites*).

En cherchant à fournir l'accès au document à un vaste public, E1 retient parfois des termes très spécifiques. D'ailleurs, la justification du choix de la plupart des noms propres et des lieux se rapporte à l'utilisateur. Rappelons que près de la moitié des DEgéo attribués pour l'ensemble des indexats proviennent des listes de E1.

#### **E1 Étang aéré v2**

Alors quelqu'un peut très bien aller chercher là-dessus (*traitement eau usée*). Et c'est un terme plus connu. Tandis que les autres (*épuration, biofiltration, étang aéré*), peuvent très bien être utilisés par les chercheurs du ministère. C'est important de souligner les deux.

<sup>9</sup> Rappelons que nous utilisons l'expression usager(s) pour désigner la dimension comprenant la prise en compte des besoins des usagers, la prise en compte d'un type particulier d'usagers et la référence à l'utilisateur en tant qu'utilisateur de la base de données.

**E1 Biogaz v1**

Alors, est-ce qu'on écrit ce que c'est ce *ISCST*? Ça pas l'air. Mais pour les chercheurs, ça devrait être suffisant. C'est tellement technique, ça prend vraiment un habitué qui va comprendre (*elle inscrit ISCST en brouillon*).

**E1 Biogaz v2**

**expé**

Tu mets les deux, *Montréal* et *CUM*.

**E1**

Quelqu'un ne pensera pas *CUM*, d'autres vont y penser tout de suite. Moi personnellement, je penserais plus à *Montréal*. Mais ceux qui sont vraiment en environnement vont plus penser territoire de la *CUM*. Il faut tenir compte de ça.

Sa prise en compte constante de l'utilisateur se manifeste par le choix du terme le plus connu, du terme employé par les spécialistes, et par le rejet de termes ne possédant pas une capacité de discrimination suffisante pour permettre le repérage des notices pertinentes et ainsi minimiser le bruit au repérage.

**E2**

De prime abord, E2 ne semble pas indexer pour un type d'utilisateur particulier, mais plusieurs de ses énoncés indiquent qu'elle n'indexe toutefois pas pour des spécialistes.

**E2 Étang aéré v2**

Des utilisateurs en environnement avec une banque en environnement, c'est sûr. Je ne pense pas à une catégorie de gens en particulier. J'aurais pu penser aux étudiants ou les industriels, non. Je ne m'arrête pas au type de consultation.

**E2 Compostage v2**

Un terme technique, on ne met pas ça, ce n'est pas nécessaire. Ce n'est pas tout le monde qui est au courant que ça s'appelle *triga* et etc. (*les différentes techniques de compostage*).

L'utilisateur n'est pas au centre des préoccupations de E2, mais sa prise en compte est principalement associée au souci de faciliter l'interprétation du contenu du document à partir de la seule lecture de la notice. Il pourrait s'agir alors d'une prise en compte de l'utilisateur en post-repérage, c'est-à-dire, une fois que l'utilisateur lit les notices retrouvées par la base de données. Ainsi, l'ordre des descripteurs et leur distribution dans les champs DE1 et DE2 sont effectués en ce sens et certains descripteurs plus génériques (*traitement eau usée, sources d'énergies*) sont retenus explicitement en fonction de l'utilisateur.

**E2 Étang aéré v2**

J'essaie le plus possible de les mettre en ordre. Un ordre qui va plus révéler de quoi parle le sujet. C'est comme s'il avait le résumé du document, mais juste des mots, et que la personne qui voit cela fasse un lien entre chaque.

**E2 Biogaz v2**

Dans DE1, l'utilisateur veut savoir tout suite à quoi s'attendre du document.

**E2 Étang aéré v2**

À ce moment-là, je pense que je me mets vraiment à la place de l'utilisateur encore plus, parce que je me dis, c'est comme si ce volume-là était déjà dans la banque et quelqu'un voudrait le rechercher. Je relis en fonction de l'utilisateur, s'il serait satisfait d'après les termes qu'il y a là.

À l'étape de révision, E2 évalue sa liste de descripteurs en fonction de la satisfaction de l'utilisateur.

**E3**

E3 dit connaître la clientèle du MENVIQ, mais ne semble pas avoir une idée précise du type d'utilisateurs auquel cette clientèle correspond.

**E3 Étang aéré v2**

C'est vague. Je pense plus aux gens qui travaillent au ministère de l'environnement, et qu'ils ont affaire aux chercheurs qui travaillent au ministère de l'environnement. Je pense que c'est plus ça que j'ai en tête. Parce que notre premier but c'est de desservir la clientèle ministérielle. Les étudiants ou monsieur-tout-le-monde, ils sont les bienvenus, mais notre but premier ce n'est pas de leur répondre.

E3 évalue plusieurs termes en fonction de la satisfaction de l'utilisateur, mais cet élément est rarement évoqué seul pour le choix d'un terme. De fait, il semble que l'importance du terme dans sa représentation du contenu du document et sa présence dans le thésaurus priment sur tout autre élément. Ainsi, la définition de la satisfaction de l'utilisateur apparaît conditionnelle à l'application de ces facteurs. Généralement, E3 mentionne l'utilisateur à l'étape de révision finale de sa liste de descripteurs.

**E3 Étang aéré v2**

Souvent en dernier quand je révise, je me dis souvent, si quelqu'un cherche un document là-dessus, pis il retrouve cela, va-t-il être content?

**E3 Étang aéré v2**

Je ne sais pas, peut-être que si *contraintes* avait été dans le thésaurus, je l'aurais peut-être mis. Dans ma tête j'aurais pensé, tiens *contraintes* on a ça je vais mettre ça. Mais vu qu'il ne l'est pas...

C'est des *contraintes*, mais dans le fond ça rapport à l'évaluation des sites. *Contraintes* c'est aussi vague que l'évaluation.... Il n'y a personne qui va chercher là dessus.

**E3 Compostage v2**

Il (*zone urbaine*) est dans le titre aussi. Sauf que... on pourrait aussi chercher d'autres genres de documents avec ce descripteur. Tandis que si on cherche des choses sur le compostage, ça rentre en plein dedans. Je pense souvent en fonction de la recherche. Je me dis que si quelqu'un cherche des choses avec ces descripteurs là, il va retrouver ça qui est en plein dedans ça va être correct, ... mais on peut chercher avec ce descripteur (*zone urbaine*) et ne pas s'attendre à trouver ce document.

Notons au sujet de ce dernier énoncé que E3 retiendra *zones urbaines* en descripteurs.

Par ailleurs, il semble que E3 oblige l'utilisateur à adapter sa stratégie de recherche en fonction de sa propre représentation du contenu en descripteurs. Ainsi, selon E3, l'utilisateur qui fait sa requête avec une expression spécifique, telle que *design des étangs, biomasse* ou *caractérisation du biogaz*, et ne retrouve aucune notice, devra alors penser à chercher à un niveau plus générique ou à consulter le thésaurus afin de sélectionner le descripteur approprié pour questionner la base de données.

**E3 Biogaz v1**

Caractérisation du biogaz. Caractérisation. Devrais-je le mettre?

**E3 Biogaz v2**

Il me restait un petit doute. *Caractérisation* occupe une bonne partie du document sauf que le but de ce document c'est une étude d'impact. J'ai résolu en me disant ça (*que le but du document est lié à une étude d'impact*), mais pas nécessairement de caractériser le biogaz. Non, je le mets pas. Si quelqu'un cherche la *caractérisation du biogaz* et ne trouve rien, il va aller voir à *biogaz* et va sortir ça et donc va l'avoir quand même.

**E3 Énergie v2**

S'il (*l'utilisateur*) va voir à *énergie de la biomasse* (*dans le thésaurus*), il va voir aussi *boise comme combustible*, il peut toujours l'essayer. De toute façon, il est là-dedans (*dans le thésaurus*) le renvoi est là.

Ceci nous laisse donc croire que E3 ne centre pas son analyse autour de l'utilisateur. En fait, lorsqu'elle mentionne cet élément à l'étape de révision finale, la satisfaction de l'utilisateur est évaluée conjointement à la pertinence des descripteurs dans leur représentation du contenu du document.

**E4**

L'ensemble des protocoles de E4 semblent indiquer que l'élément utilisateur n'est, à toute fin utile, pas pris en compte pour le choix de ses termes et descripteurs. En effet, le protocole du document *Énergie* ne comporte aucune mention de l'utilisateur et celui du document *Compostage* ne contient qu'une seule mention. La plupart des verbalisations qui concernent directement l'utilisateur ont été provoquées par une question explicite de

l'expérimentateur puisque E4 n'en faisant aucune mention spontanée. Ainsi, il semble normal que E4 n'ait aucun type d'usager particulier en tête lorsqu'elle indexe.

#### **E4 Étang aéré v2**

**expé**

As-tu un usager dans la tête quand tu choisis les mots?

**E4**

Non. Pas nécessairement. Bien que j'avoue que souvent je me suis dit, peut-être pas dans celui-là (*ce document*) mais pour un autre, bien oui je vais mettre *circulation* avec *avion*, ça va me donner quelque chose. Si je fais la recherche... Des fois je pense, mais ce n'est pas là tout le temps.

#### **E4 Biogaz v2**

C'est une chose dont on entend beaucoup parler (*effet de serre*), beaucoup d'écrits là-dessus. C'est peut-être une pensée pour les chercheurs. Tu ne passes pas à côté d'un document qui parle d'*effet de serre* sans le souligner.

En fait, les rares mentions de l'usager sont liées à d'autres éléments de connaissance privilégiés par E4, tels que l'actualité pour le terme *effet de serre* dans l'énoncé rapporté ci-dessus.

Par ailleurs, E4 souligne en entrevue qu'elle n'indexe pas pour l'usager puisqu'il est difficile de prédire ses besoins et questions *a priori* étant donné qu'ils sont de nature et de niveaux très variés. De plus, les indexeurs étant rarement en contact direct avec les usagers, il leur est difficile de prévoir leurs requêtes.

### **2.3.4 Éléments de connaissance supplémentaires**

Tout au long des analyses de protocoles, d'autres éléments de connaissance, dont nous n'avions pas anticipé l'importance, se sont révélés décisifs dans certains cas comme facteur explicatif des divergences interindexeurs. Comme pour les éléments déjà analysés, il est rare que l'utilisation d'un de ces éléments s'effectue de manière isolée. Pour simplifier la distinction interindexeurs, ils sont présentés séparément à cette étape-ci et seront intégrés dans la caractérisation globale d'indexation de chaque indexeur présentée à la section hiérarchisation de l'utilisation des éléments de connaissance (section 2.5).

La norme ISO (1985) ne fait pas référence directement à ces éléments, mais on peut supposer que les indexeurs restent à l'affût de tout aspect présentant un intérêt pour la caractérisation du document et une utilité potentielle pour les usagers, ce que souligne la

norme: "Aussi, il est recommandé, par exemple, aux indexeurs de littérature scientifique et technique, de ne pas négliger les autres facettes d'un sujet, par exemple son aspect social ou économique" (ISO, 1985, p.3).

Nous débutons cette section par une définition globale de chaque élément pour ensuite caractériser l'utilisation particulière qui en est faite par les indexeurs. Soulignons que les définitions que nous en donnons ont été élaborées à partir de l'analyse des verbalisations des indexeurs. Cinq éléments de connaissance supplémentaires ont été identifiés, lesquels ont été nommés actualité, nouveauté, rareté, capacité de discrimination et propos central.

Parmi ces éléments, l'identification du caractère de nouveauté, de rareté et la capacité de discrimination d'un terme se rapportent à la connaissance du domaine et du fonds documentaire.

Le caractère de *nouveauté* s'apparente aux notions qui sont nouvellement utilisées en environnement, faisant ainsi référence à l'évolution du vocabulaire de ce domaine. Les indexeurs doivent connaître relativement bien les problématiques et les sujets d'un domaine pour être en mesure d'identifier les nouvelles notions.

La *rareté* est associée aux notions qui sont rarement abordées dans les documents traités par un centre documentaire particulier, en l'occurrence les documents de la base de données Envirodoq. Il peut s'agir de notions bien connues dans le domaine dont les documents de ce centre ne traitent que rarement en comparaison à un terme qui est nouvellement utilisé. Il devient alors intéressant de caractériser le document par cette notion permettant ainsi de le distinguer des autres documents appartenant au même fonds.

Ceci rejoint l'élément de la *capacité de discrimination* d'un terme. Dans un premier temps, il peut s'agir pour l'indexeur de vouloir souligner un aspect intéressant du document, qu'il soit question d'une notion rare ou nouvelle, permettant de singulariser le document par rapport aux autres documents traitant de sujets similaires. Cependant, bien qu'elle puisse être associée à la rareté et à la nouveauté, la capacité de discrimination en est distincte. En effet, une notion peut-être retenue pour sa capacité de discrimination sans pour autant être

rare, nouvelle ou d'actualité. Également, cet élément de connaissance peut être étroitement lié à la prise en compte de l'utilisateur.

On entend par *actualité* toute notion à laquelle l'indexeur fait référence en rapport à l'information véhiculée par les médias ainsi que les notions plus "à la mode" dans le domaine de l'environnement. Évidemment, le caractère d'actualité d'un terme n'est pas permanent et correspond à une période donnée comme le soulignent d'ailleurs plusieurs indexeurs.

Finalement, nous avons ajouté un élément de connaissance, soit le *propos central*. En fait, cet élément correspond à un trait particulier du processus d'un indexeur (E3). Mais étant donné que son analyse est vraiment centrée autour du propos central et que pour cet indexeur, le *satisficing* est synonyme de la quasi unique représentation du propos premier du document, nous nous devons d'ajouter un élément qui permette de faire ressortir cette particularité.

Nous avons également identifié une stratégie personnelle liée à l'ordonnement des descripteurs. Il ne s'agit pas de l'application d'un élément de connaissance, mais comme cela permet de mieux caractériser le processus de trois des quatre indexeurs, il en est fait mention à la présente section.

Examinons maintenant la façon dont la prise en compte de ces éléments s'articule à l'intérieur du processus d'indexation de chaque indexeur.

## E1

Afin d'être en mesure de comprendre les décisions de E1 et caractériser son processus d'indexation par rapport aux autres indexeurs, on doit s'attarder à sa prise en compte de plusieurs éléments qui jouent un rôle important pour le choix de ses termes, soit la capacité de discrimination, la rareté et l'actualité.

Premièrement, le processus de E1 peut être caractérisé par une prise en compte de la capacité de discrimination des termes afin d'évaluer leur pertinence. Ainsi, E1 retient des

termes qui permettent de situer un document parmi l'ensemble des autres documents semblables de la base de données, sans que le document en traite de manière substantielle; ce qui renvoie à sa connaissance du fonds documentaire. Donc, une requête avec ces termes ne permettra pas de réduire le repérage aux seuls documents très pertinents, mais indiquera plutôt un aspect particulier du document. Ce facteur motive le choix de termes, tels que *ISCST*, *informatique*, *logiciel*, et les différents *sites* du document *Étang aéré*, qui répondront à une demande très spécifique de la part d'utilisateurs spécialistes.

#### **E1 Biogaz v2**

Parce que de plus en plus on voit des *logiciels* spécialisés dans n'importe quel domaine, mais en environnement entre autres, pour aider à effectuer telle tâche, telle évaluation. J'imagine que s'il y a des gens qui peuvent retracer cela directement, ça leur évite de se taper plusieurs études.

De même, la prise en compte de la capacité de discrimination entraîne le rejet de termes couverts raisonnablement, mais qui ne permettent pas de particulariser un document parmi l'ensemble des documents de la base puisqu'un nombre considérable de documents en environnement traitent de ces notions. Ainsi, des termes, tels que *évaluation*, *échantillonnage*, *contraintes*, *usine*, *implantation*, *expérimentation*, etc. sont rejetés puisqu'ils causeraient trop de bruit au repérage. Notons que plusieurs de ces termes sont retenus par d'autres indexeurs.

#### **E1 Étang aéré v2**

Pourquoi mettre des mots qui n'ont pas vraiment leur utilité. Cette base-là s'adresse à des chercheurs en environnement. Alors pourquoi aller mettre *usine*. Pourquoi mettre *dézonage*. C'est des choses courantes qui arrivent dans les études environnementales.

#### **E1 Étang aéré v2**

Pour expliquer son document, il doit y avoir un chapitre qui explique justement le titre et l'étude en question.... Ça se fait par *évaluation*, par *échantillonnage*, par *expérimentation*, ensuite il y a une *évaluation* et puis il y a toujours des *contraintes* de différentes formes. S'ils n'avaient pas été dans la table des matières, j'aurais passé à côté ....

Certains de ces termes sont également retenus à cause de leur caractère d'actualité ou de "mode" et sont jugés susceptibles d'intéresser l'utilisateur.

#### **E1 Biogaz v2**

(*Carrière Miron*) on en parle tout de suite au début du document même si ce n'est pas très long. Justement, par ta connaissance de l'actualité, ça fait longtemps qu'on en entend parler, et on va en entendre parler encore longtemps.

Même si les termes choisis pour leur capacité de discrimination ne sont pas utilisés par l'utilisateur, mais sont compris dans la notice repérée, selon E1, ils pourront néanmoins aider l'utilisateur à choisir les notices les plus pertinentes parmi l'ensemble des notices repérées. Il s'agit alors d'une prise en compte de l'utilisateur en post-repérage, c'est-à-dire une fois que l'utilisateur a repéré un ensemble de notices.

#### **E1 Biogaz v2**

expé

Celui-là dépendrait de quoi? C'est quoi son lien, son statut...

**E1**

Ça fait partie de la *modélisation*, l'aspect *informatique*. Comme modélisation je ne peux pas le mettre dans les principaux... Il n'était pas dans le thésaurus.... Il me semble que quelqu'un qui voit juste *informatique* tout seul va se dire ce n'est pas le document en tant que tel qui traite d'informatique mais il y a un aspect dans l'évaluation environnementale qui est informatisé.

#### **E1 Biogaz v2**

(ISCST) encore une fois. On parle beaucoup de modélisation. Mais un autre que l'on fait ressortir, c'est lui (*le modèle ISCST*). Alors moi, je me dis que si le chercheur voit que dans la notice, il est identifié ce modèle. C'est sûr qu'on va parler de d'autres modèles ou bien de modélisation de d'autres types.

À travers cette évaluation multiple de la capacité de discrimination des termes, on observe le souci quasi permanent pour l'utilisateur sous-jacent à ses décisions. Cette constatation fait apparaître le rôle prédominant des dimensions usagers et capacité de discrimination dans le processus d'indexation de E1 et permet de la caractériser par rapport aux processus des autres indexeurs.

Par ailleurs, E1 tient à représenter les notions qui sont plus rarement abordées dans les documents en environnement. La rareté d'un terme permet également de discriminer un nombre restreint de documents de la base, mais se distingue de la capacité de discrimination utilisée par E1 puisque ces termes ne semblent pas choisis en fonction d'un groupe particulier d'utilisateurs. En fait, le choix d'un terme jugé rare implique un potentiel de discrimination, mais un terme choisi pour sa capacité de discrimination ne renferme pas nécessairement l'élément de rareté. Notons que cette distinction est subtile et que ces deux éléments peuvent être confondus par moment. Dans le cas présent, on se base sur les verbalisations de E1 pour établir la différence.

#### **E1 Énergie v2**

expé

Pourquoi ce Ah! recommandations?

**E1**

Dans les autres documents, ce n'était pas spécifié un chapitre sur les *recommandations*. Celui-là le fait alors c'est important de le mentionner, comme ça ne se présente pas partout.

**E1 Biogaz v2**

Ils ont même fait une abréviation. Si on se pose la question en regardant juste les DE, en prenant du recul, il doit y en avoir plein de document qui ont comme DE *élimination déchet, enfouissement sanitaire*. Cela ne me dit pas que c'est une étude sur la *qualité de l'air* où on parle beaucoup des *COV (composés organiques volatils)*. Alors ce serait important de le mettre pour différencier cette étude-là des autres qui traitent de *l'élimination des déchets*.

**E1 Compostage**

(*Écrit coûts/bénéfices, analyse*) comme j'ai dit, ce n'est pas toutes les études qui mentionnent le commerce, l'économie, les coûts...

Soulignons que les termes *recommandations* et *revue de littérature* (retenus pour le document *Biogaz*) n'ont pu être traduits et ont été abandonnés.

**E2**

E2 ne semble pas se préoccuper de l'un ou l'autre des éléments de connaissance supplémentaires identifiés. En effet, on ne trouve pas de trace de ces éléments dans ses protocoles. Cependant, E2 se démarque des autres indexeurs par la recherche des différents aspects ou facettes présentés dans un document. Elle réfère à ce principe par les expressions "entourer le sujet", "faire le tour du sujet" lors de sa révision de sa liste de termes brouillon et sa liste de descripteurs.

**E2 Étang aéré v1**

Bon! Là, je résume en gros, si j'ai bien noté. (*Relit ses descripteurs*) L'emplacement, étangs aérés, traitement eaux usées, municipalité, évaluation. DE2: répercussion, station d'épuration, coûts. Là, on voit réellement. Je pense que j'ai fait pas mal le tour de ce document-là!

**E2 Compostage v2**

Je n'ai pas l'impression d'avoir fait le tour du document. Je suis plus satisfaite des autres documents (*Étang aéré* et *Biogaz*). Je suis certaine que je n'en ai pas assez, j'en aurais mis plus encore.

**E2 Énergie v1**

Je pense que j'ai fait pas mal le tour de la question. Il y a quand même des choses que je ne suis pas capable de satisfaire, genre *équipement* ou quelque chose comme ça.

En fait, il s'agit d'une application particulière de l'exhaustivité, visant à n'omettre aucune notion importante du document pouvant intéresser l'utilisateur. Ainsi, les différents aspects ou facettes ne sont pas prédéterminés, telle l'application d'une grille, mais plutôt identifiés au cours de l'analyse du document et correspondent aux notions couvertes suffisamment par le document justifiant ainsi leur sélection.

**E2 Étang aéré v2**

Pour moi, cerner le sujet, entourer le sujet, c'est comme une révision, si ça de l'allure. Parce qu'au début on ne connaît pas le document. On sort des choses, on en prend connaissance au fur et à mesure. En dernier, on relit ça, on se met à la place de l'utilisateur. Et je me dis, ai-je vraiment tout sorti ce qu'il y avait d'essentiel par les termes. On cherche à le mettre le plus accessible possible. Ça veut dire avoir les bons descripteurs.

En plus des thèmes principaux, E2 s'attarde habituellement aux facettes suivantes: résultat, lieu (qui répond à une politique du MENVIQ), but du document, plus certains aspects reliés au contenu particulier de chaque document.

**E2 Étang aéré v2**

*Traitement eau usée* parce que c'est le but des sites d'implantation. C'est une raison du document. Ça spécifie plus mes *sites*. De plus, ce descripteur aide l'utilisateur qui ne sait pas que *étang aéré* implique le *traitement des eaux usées*.

**E2 Étang aéré v2**

C'est des *sites d'implantation* à quel endroit, c'est au point de vue *municipalité*. Ce n'est pas un peu partout, c'est vraiment un endroit spécifique.

**E2 Biogaz v2**

(*Par rapport au terme échantillonnage*) Là, j'allais voir s'il y a vraiment des résultats. C'était le cas, les données sont dans le document. On le met quand il y a des prélèvements et que les tableaux sont là, tel jour telle heure, les résultats. Mais il faut qu'ils en parlent vraiment, sinon on ne le mentionne pas.

**E2 Biogaz v2**

Le fait que ça se passe à *Montréal*, ici, c'est pas à l'autre bout du monde c'est plus intéressant.

**E2 Énergie v2**

J'accroche parce que c'est quand même important. Milieu humain, physique. Ils décrivent comment, dans quel milieu sont ces problèmes, les besoins d'énergies et tout ça. C'est quand même une description de l'endroit où ça se passe. Habituellement, c'est important de se situer, de savoir dans quel milieu ça se passe.

**E2 Énergie v2**

Je trouvais important en DEI de situer tout de suite l'utilisateur, où ça se passait ces énergies renouvelables, en *pays en voie de développement*.

Notons toutefois que la satisfaction de l'utilisateur est souvent prise en considération pour évaluer la pertinence de représenter un des aspects.

**E2 Compostage v2**

*Agriculture*, ça me pesait. Il ne parle pas vraiment d'*agriculture*, mais le but du *compostage* et ces *boues* c'est en fonction de l'*agriculture*. Je me demandais si j'allais le mettre quand même. C'est dans le but de, mais si on met *agriculture* et un utilisateur veut avoir des documents sur l'*agriculture* et on lui sort un document sur la gestion des *boues résiduelles urbaines*. On est pris entre deux idées. Si je ne l'ai pas mis c'est parce que je me suis mise à la place de l'utilisateur qui allait trouver ça.

Soulignons que ce processus particulier de E2 sera dorénavant associé directement à son application de l'exhaustivité servant ainsi à le qualifier de manière plus précise. En ce

sens, nous n'ajoutons pas un élément de connaissance supplémentaire pour rendre compte de ce comportement bien qu'il permette de mieux caractériser E2 par rapport à ses pairs.

Finalement, rappelons que E2 tente de placer ses descripteurs en un ordre qui permette à l'usager d'avoir une idée des relations unissant les descripteurs à partir de la seule lecture de la notice, et ce, malgré le fait qu'avec un langage post-coordonné, il ne peut y avoir d'association explicite entre les différents descripteurs d'une notice.

### E3

Le processus d'indexation de E3 peut être caractérisé par le souci constant de s'en tenir presque exclusivement à la représentation du propos premier du document qu'elle recherche fréquemment par la dimension "but du document". La prise en compte de cet élément permet d'expliquer un grand nombre de choix et de rejets de termes et permet également de distinguer le processus de E3 de celui des autres indexeurs.

#### **E3 Biogaz v2**

Le but du document... s'il n'y avait pas de *biogaz*, on ne ferait pas ce document.

#### **E3 Biogaz v2**

(E3 recherche la mention du but de l'étude). Des fois dans les lettres des rapports d'étude de répercussions environnementales, ils disent dans le but de...

L'énoncé suivant rend compte du cœur de la procédure d'indexation de E3. Le terme évalué doit se rapporter directement au sujet premier du document et non être relié indirectement à ce dernier.

#### **E3 Biogaz v2**

*Échantillonnage*, ça rapport à la *caractérisation*. Quand tu veux caractériser, t'échantillonnes pour savoir de quoi est composé.... On peut faire de *l'échantillonnage* mais c'est secondaire, c'est dans le but de nous permettre d'évaluer les impacts. Je l'ai laissé tomber. Le document n'est pas sur *l'échantillonnage*.

Notons que cette préoccupation s'apparente à l'exactitude tout en étant distincte, car il ne s'agit pas uniquement de déterminer si le document traite suffisamment de la notion ou si le sens du terme est conservé une fois traduit en descripteur, mais plutôt si le terme représente l'un des aspects essentiels du document. Par extension, il est possible de qualifier le niveau d'indexation de E3 de superficiel.

Par ailleurs, E3 est le seul indexeur qui ordonne ses descripteurs en un énoncé significatif. Il ne s'agit pas d'une démarche en fonction de l'utilisateur, contrairement à E2, mais plutôt d'une façon de résumer le sens du document et de s'assurer de la pertinence du produit final.

**E3 Étang aéré v1**

*(E3 fait sa phrase)* Évaluation des emplacements pour d'étangs aérés pour le traitement des eaux usées à Piedmont.

**E3 Étang aéré v2**

Souvent j'essaie de me faire une phrase. Le document porte sur l'évaluation des emplacements des étangs aérés. J'essaies de me faire une phrase qui résume un peu le sens du document. Souvent, c'est ça dans ma tête.

**E3 Biogaz v1**

*(E3 fait sa phrase)* C'est une technologie pour l'élimination des biogaz sur un site d'enfouissement sanitaire. On évalue les répercussions environnementales. Les données sur la qualité de l'air.

**E3 Biogaz v2**

La phrase des fois peut ne pas correspondre (*à l'ordre des DE*) ça dépend du document. Mais ça correspond souvent.

Toutefois, l'introduction d'un nouveau thésaurus change légèrement les habitudes de E3 et la forme composée de la majorité des descripteurs d'Infoterra ne permet pas de combiner les descripteurs les uns à la suite des autres en un énoncé significatif.

**E3 Compostage v2**

*(Débute l'écriture de ses DE1)* j'étais désorientée.

**expé**

Est-ce qu'il y a une importance dans l'ordre que tu les mets?

**E3**

Peut-être à cause du titre, non même pas.

En dernier lieu, E3 semble porter une attention aux notions rarement abordées dans les documents de la collection du MENVIQ.

**E3 Biogaz v2**

*Bruit.* C'est un descripteur que l'on n'emploie pas souvent. On n'a pas beaucoup d'études qui parlent de la pollution par le bruit. L'utilisateur sera intéressé par ce document. Ma connaissance de la collection a jouée ici. J'aurais pu ne pas le mettre, mais cela en fait un de plus et c'est mieux que de ne pas l'avoir.

E4

E4 est réceptive au caractère d'actualité d'un terme, de rareté et de nouveauté.

**E4 Étang aéré v2**

*Design*, c'est un mot qu'on emploie et qu'on utilise et qu'on voit beaucoup. Par contre, le *design d'étang*, moi, c'était la première fois que je voyais un document avec ce terme.

**E4 Biogaz v2**

(*À propos du terme cogénération*) comme tantôt, c'est parce que ça se parle beaucoup. Et quand un terme vient à la mode comme tu ne le retrouvais pas avant parce qu'il n'y avait rien d'écrit ou parce qu'il n'existait pas, donc quand tu le rencontres tu le mets.

Notons qu'elle retient le terme *cogénération* en candidat, même si elle n'en connaît pas la signification.

Par ailleurs, un des points qui différencie l'approche de E4 de celles des autres indexeurs concerne son interprétation de la représentativité des termes pour un document. Concrètement, E4 évalue parfois l'importance d'un terme ou d'une notion en fonction de sa connaissance générale du problème ou de la situation traitée dans le document plutôt que de limiter son évaluation à la représentativité du terme pour le contenu du document. Il s'ensuit qu'elle retient des termes qu'aucun autre indexeur ne jugerait adéquats, tels que *capacités optimales du sol, sécheresse, vent, précipitations, climatologie*, etc., pour le document *Énergie et effet serre, méthane, cogénération*, pour le document *Biogaz*.

**E4 Énergie v2**

*Capacités optimales du sol*, ça je le garde parce que dans ce coin-là, c'est important de tirer du sol le plus qu'il peut te donner.

Elle conserve ainsi plusieurs termes, malgré le fait que le sujet du document ne porte pas sur ces aspects, c'est-à-dire que ces aspects n'occupent qu'une mince portion du texte, parfois même une seule phrase.

Nous n'avons pas attribué un élément de connaissance particulier à ce comportement unique, mais nous supposons qu'il peut être associé à sa prise en compte de l'exhaustivité.

De plus, cette interprétation particulière de l'importance des termes du document liée à son souci pour l'exhaustivité, entraîne E4 à recourir à la notion de coextensivité beaucoup plus fréquemment que les autres indexeurs. Ce principe consiste à sélectionner un descripteur dont une partie seulement correspond au concept que l'on veut exprimer, ou à retenir deux descripteurs afin de représenter complètement un concept. Par exemple, la

sélection du descripteur *terres agricoles* pour représenter le concept de *valorisation agricole* et l'utilisation des deux descripteurs *puits* et *contamination* pour rendre la notion de *contamination des puits*.

Évidemment, l'introduction du thésaurus Infoterra comprenant des descripteurs composés, obligera les indexeurs à recourir à la coextensivité, mais E4 en fait une utilisation relativement intense. Pour l'indexation des deux documents avec le thésaurus Envirodoq, il arrive également qu'elle recoure à ce principe, que l'on explique par son souci marqué pour l'exhaustivité.

#### **E4 Biogaz v2**

(**Circulation, avion**) Je les ai mis comme si c'était vraiment juste un terme (*E4 retient ces deux descripteurs pour rendre l'expression transport aérien*).

Finalement, E4 souligne l'attribution d'un ordre d'importance à ses descripteurs, les premiers de la liste étant les plus importants, et ce également à l'intérieur d'un même champ (DE1 ou DE2). Bien que cette distinction corresponde en partie à la procédure standard<sup>10</sup>, nous mentionnons ce détail puisque E4 elle-même s'attarde à cette dimension de son processus et que les indexeurs E2 et E3 rendent compte également de leur approche personnelle concernant l'ordre des descripteurs.

#### **E4 Étang aéré v2**

Quand je les ai écrits, *étang aéré, emplacement, évaluation*, dans ma tête, il y avait une gradation. J'ai mis le plus important en premier et les autres sont comme un peu plus secondaires.... Mais *usine d'épuration (DE2)* se rattachait directement à *traitement eau usée (DE1)*, c'est lui que j'ai mis en premier (*dans DE2*).

#### **E4 Étang aéré v2**

Mais dans le fond, ça n'a pas d'importance que je fasse cela ou pas, c'est juste pour moi. Même les principaux et secondaires sont tous interrogeables au même niveau, ça n'a pas d'importance. Ça ne dérange rien leur place dans la grille. Je pense que c'est une idée à moi effectivement, de garder une certaine harmonie dans les DE.

#### **E4 Biogaz v2**

Quoique si je regarde ma feuille, j'ai encore mis d'un bord ceux que je considérais comme plus principaux. On dirait que ceux-là me viennent facilement, je les repère facilement.

#### **E4 Biogaz v2**

Les quatre premiers sont peut-être plus généraux. Après cela je rentre dans... Si on fait des petites bulles, les petites bulles qui viendraient les dernières autour.

---

<sup>10</sup> La procédure souligne une distinction relativement claire entre les champs des DE1 et DE2, mais pas entre les descripteurs d'un même champ.

Rappelons en dernier lieu, que cet ordre de priorité a été bousculé avec l'introduction du nouveau thésaurus.

## 2.4 Comparaison interindexeurs pour l'utilisation des éléments de connaissance.

Les différentes prises en compte des éléments de connaissance étant identifiées, il est important de cerner ce que ces distinctions et ressemblances interindexeurs ont permis de mettre à jour pour expliquer les variations de choix de termes.

Tout d'abord, examinons le Tableau XVI qui permet de comparer les différents ensembles d'éléments privilégiés par chaque indexeur. Soulignons que cette vue d'ensemble ne révèle pas la nature de la prise en compte d'un élément et s'avère donc une représentation simplificatrice, mais elle permet de rassembler en un seul tableau une première comparaison interindexeurs systématique.

Tableau XVI. Ensemble des éléments de connaissance privilégiés par chaque indexeur.

ÉLÉMENTS DE CONNAISSANCE	INDEXEURS			
	E1	E2	E3	E4
• Exhaustivité		√		√
• Précision			√	
• Exactitude	√	√	√	
• Rareté	√		√	√
• Actualité	√			√
• Nouveauté				√
• Capacité de discrimination	√			
• Propos central			√	
• Usager	√	√		
<b>Ordre des descripteurs:</b>				
• Indiquer relations entre les DE		√		
• Énoncé significatif			√	
• Ordre d'importance				√

Ainsi, un crochet indique que l'indexeur utilise cet élément. Cependant, il est possible que ce recours ne corresponde pas exactement à la prescription de la norme ISO et reflète une appréhension plus personnelle, ou encore, il se peut que l'indexeur tienne

compte d'un seul aspect de l'élément dans son application. Par exemple, un indexeur peut appliquer le critère d'exhaustivité en identifiant l'ensemble des notions présentées dans un document, mais ne pas tenir compte de la valeur potentielle des termes pour les usagers (tel que recommandé par la norme pour l'application de l'exhaustivité). Une case vide indique que l'indexeur n'utilise pas cet élément de connaissance ou que les rares mentions qui y sont faites démontrent que le choix de termes ne repose pas sur cet élément, ce qui lui confère alors une influence négligeable dans le processus d'indexation de cet indexeur. Finalement, nous avons ajouté une section intitulée, ordre des descripteurs, qui permet de distinguer les indexeurs en fonction de leur façon d'ordonner les descripteurs de leurs listes, ce qui témoigne d'une stratégie personnelle.

Un premier examen du Tableau XVI révèle que l'exhaustivité est prise en compte par deux indexeurs, un seul indexeur mise sur la précision alors que trois indexeurs privilégient l'exactitude. L'utilisateur est considéré par deux indexeurs. Parmi les éléments de connaissance supplémentaires fréquemment évoqués par les indexeurs, la rareté est prise en compte par trois indexeurs, deux indexeurs s'attardent à la dimension d'actualité. Par ailleurs, on remarque que chaque indexeur se distingue par une prise en compte unique d'un ou plusieurs éléments permettant de mieux définir sa stratégie. Ainsi, E1 se distingue principalement par l'élément capacité de discrimination, E3 par le propos central et E4 par l'élément de nouveauté. Rappelons que E2 se distingue par le souci de représenter les différentes facettes, ce que nous avons inclus dans l'exhaustivité.

Outre le fait que chaque indexeur peut être caractérisé par l'utilisation d'une combinaison différente d'éléments de connaissance, qu'est-ce que l'utilisation particulière de ces éléments nous révèle sur les différences de choix de termes?

Comparons l'application de l'exhaustivité par les deux indexeurs les plus exhaustifs, E2 et E4. Alors que E2 se limite à l'utilisation d'un descripteur pour chaque facette du document en misant sur son exactitude, E4 retient la plupart des termes présents dans les parties les plus informatives et ceux mis en évidence par une typographie particulière dans le corps du texte, sans évaluer la valeur réelle de ces termes pour la représentation du contenu du document. À titre d'exemple, comparons leur évaluation de la notion de *règlement* pour le document *Biogaz*.

**E4 Biogaz v1**

Comparer les émissions et les retombées atmosphériques prévues avec le Cadre juridique en vigueur ainsi qu'avec les plus récentes exigences canadiennes et californiennes... Ah, là, y aurait peut-être législation?

**E4 Biogaz v2**

Pour couvrir l'aspect de normes, règlements, lois en vigueur au Canada, un peu partout. Le terme c'est législation.

Dans ce dernier énoncé, E4 identifie une notion et retient un terme sans questionner sa pertinence. Elle représentera d'ailleurs cette notion par les descripteurs *règlement* et *législation* sur sa liste finale.

**E2 Biogaz v1**

L'impact sur la qualité de l'air est évalué de deux façons. La première consiste à comparer les émissions aux exigences du règlement...

**E2 Biogaz v2**

*Règlement*, là j'avoue que j'ai pensé qu'il y avait un petit côté droit, règlement, législation. Mais ça dépend à quel niveau il est développé. Là je me suis dit qu'il faudrait que j'aille vérifier dans la table des matières si le côté *règlement* est vraiment développé.

Finalement, E2 ne retient pas ce terme. Bien que la raison ne soit pas évoquée explicitement, on peut supposer qu'un aspect qui ne couvre que deux pages du document n'est pas jugé pertinent.

Un point majeur permet de distinguer ces deux indexeurs dans leur application de l'exhaustivité. Pour E2, l'exhaustivité semble tributaire de l'exactitude, alors que E4 applique l'exhaustivité de manière inconditionnelle. Ceci démontre que pour pouvoir expliquer les variations de choix de termes, l'on doit nécessairement considérer la combinaison d'éléments qui entrent en jeu lors du traitement d'un terme.

À ce propos, comparons l'évaluation du terme *logiciel* par deux indexeurs qui tiennent compte de l'exactitude, mais qui n'évaluent pas ce terme de la même façon.

**E2 Biogaz v1**

La seconde évaluation de l'impact sur la qualité de l'air est effectuée en modélisant la dispersion des polluants dans l'atmosphère à l'aide du logiciel.

**E2 Biogaz v2**

Je me demande s'il n'y avait pas un *système, logiciel*, une étude sur un modèle qui avait été faite ou quelque chose du genre. Et encore, si cela avait été développé dans le document.

L'argument évoqué par E2 pour évaluer la pertinence de ce terme se rapporte à l'exactitude, en fonction de l'ampleur du traitement. Il ne semble pas que cet aspect soit suffisamment couvert par le document puisque E2 ne le retient pas sur sa liste brouillon.

Par contre, E1 qui tient compte régulièrement de l'exactitude, n'évoque pas ce critère pour le traitement du terme *logiciel*, mais se réfère plutôt aux éléments d'actualité et d'usager pour justifier son choix.

#### **E1 Biogaz v2**

De plus en plus on voit des *logiciels* spécialisés dans n'importe quel domaine, mais en environnement entre autre, pour aider à effectuer telle tâche, telle évaluation. J'imagine que s'il y a des gens qui peuvent retracer cela directement, ça leur évite de se taper plusieurs études.

Cependant, tout comme E2, E1 se réfère à l'exactitude pour rejeter la notion de *règlement*.

#### **E1 Biogaz v2**

(*Évalue la notion de cadre juridique*) Mais ç'aurait pu être bien important. Mais c'est court, c'est tout court.

Peut-on supposer que l'exactitude est de première importance pour E1 lorsqu'il s'agit d'un terme qui n'est pas d'actualité dans le domaine de l'environnement, ce qui le rend moins susceptible d'intéresser un type précis d'usagers? Ceci sous-tendrait alors que E1 accorde possiblement la priorité à l'actualité et à l'usager, contrairement à E2. Ceci expliquerait en retour leur évaluation différente des termes reliés à l'actualité en environnement et pouvant répondre à une clientèle précise.

La prise en compte de la précision par E3 permet de caractériser la nature de plusieurs de ses choix et conséquemment, d'expliquer qu'elle évalue certains termes différemment des trois autres indexeurs qui se réfèrent peu à ce critère. Les énoncés suivants, extraits de l'évaluation comparée, font ressortir son souci de la précision, tout en soulignant son point de vue sur l'importance que les autres indexeurs accordent à ce critère.

#### **E3 Évaluation comparée/ Commentaires sur l'indexation du document Biogaz par E2.**

Ici (sur la liste de E2) *élimination déchet*, dans le fond c'est un peu comme *enfouissement sanitaire*. J'ai mis *enfouissement* parce que je trouvais que c'était plus précis qu'*élimination déchet*. Mais tous ces problèmes sont causés parce qu'on veut éliminer les déchets et je pense que cette personne a été portée à le mettre parce que c'était dans le titre.

**E3 Évaluation comparée/ Commentaires sur l'indexation du document Biogaz par E4.**

*Dépotoir, c'est moins précis qu'enfouissement sanitaire parce que tu peux avoir des dépotoirs à ciel ouvert, sauf qu'elle n'a pas pensé d'aller plus précis ou cela la satisfaisait d'aller large comme ça.*

*Mais moi, je préfère aller toujours au plus précis. Parce que c'est l'enfouissement sanitaire qui produit le biogaz.*

Les remarques de E3 sur une appréhension différente de la précision par les autres indexeurs s'avèrent juste puisque E2 tient à une représentation exacte mais pas nécessairement précise, et E4 fait un relevé exhaustif des notions présentées dans les parties les plus informatives sans évaluer leur exactitude ou leur précision.

Cependant, pour expliquer les différences principales entre E3 et les autres indexeurs, l'on doit s'attarder à la combinaison d'éléments utilisés par E3 et pas uniquement au critère de précision. Ainsi, la combinaison des éléments précision, exactitude et propos central semblent être à la base de la majorité des choix de E3. Par exemple, ceci permet d'expliquer que E3 ne retient pas des termes tels que *coût*, *station d'épuration* et les termes géographiques *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse* pour le document *Étang aéré*, qui répondent aux critères de précision ou d'exactitude mais qui ne représentent pas le propos central.

Par ailleurs, E1 et E2 se distinguent par la nature de leur prise en compte de l'utilisateur. Alors que E1 retient des termes spécifiquement pour répondre aux besoins potentiels d'utilisateurs spécialistes et d'autres termes pour répondre à des utilisateurs grand public, E2 évalue généralement l'ensemble des termes de sa liste en fonction de la satisfaction de l'utilisateur, une fois sa liste complétée. En comparaison à E2, la prise en compte de l'utilisateur a une influence plus directe sur les choix de termes de E1.

La principale différence interindexeurs que l'on note pour l'utilisation des éléments supplémentaires réside dans le nombre de termes retenus en fonction de chaque élément. Ainsi, bien que plusieurs indexeurs s'attardent aux notions rarement rencontrées, cet élément n'explique le choix que d'un terme des listes de E3 (*bruit*). E1 souligne le caractère inusité de plusieurs notions mais n'en retient qu'une seule en descripteur (*coûts/bénéfices*, *analyse*) et en conséquence, seul le choix de ce descripteur est lié directement à cet élément. De même, l'élément de nouveauté rend compte du choix d'un seul terme (*design*) des listes de E4. La dimension solution est associée au choix de deux candidats

descripteurs (*torchère, ISCST*) des listes de E1. Par ailleurs, l'aspect d'actualité d'un élément est évoqué plus fréquemment par les indexeurs et rend compte d'un plus grand nombre de termes retenus, particulièrement parmi les listes de E4 (*effet serre, méthane, cogénération, biogaz*).

Il est probable que le peu de références à ce groupe d'éléments, et particulièrement aux éléments de rareté et de nouveauté, témoignent en réalité du rythme de l'évolution du vocabulaire dans le domaine de l'environnement, c'est-à-dire que les notions nouvelles ou rarement abordées n'apparaissent pas dans chaque document. Ainsi, l'on ne doit pas considérer ces éléments comme ayant une influence négligeable *a priori* sur le processus d'indexation des indexeurs uniquement en fonction de la fréquence d'utilisation. À notre avis, pour pouvoir saisir leur impact réel, l'on doit plutôt chercher à identifier l'importance qu'un indexeur accorde à chacun de ces éléments en rapport à l'ensemble des éléments de connaissance utilisés.

En effet, puisque la combinaison d'éléments semble révéler de manière plus claire l'influence de chaque élément dans le processus global de l'indexation, il semble pertinent d'examiner en contexte les divers éléments afin d'identifier si leur impact sur le choix de termes n'est que ponctuel et négligeable ou s'il dirige un grand nombre de décisions prises par l'indexeur.

En somme, l'examen de l'utilisation des éléments de connaissance a permis de cerner certaines différences interindexeurs, mais cela a également fait ressortir l'importance de considérer la dynamique interéléments afin de comprendre l'influence réelle de chaque élément ou d'une combinaison d'éléments sur la résolution de l'indexation puisque de toute évidence, les éléments de connaissance ne sont pas appliqués successivement pour évaluer un terme ou une notion, mais plutôt de manière parallèle et combinée.

Ainsi, afin de caractériser les indexeurs plus distinctement, nous allons tenter de distribuer les éléments de connaissance utilisés par chaque indexeur selon un ordre hiérarchique. En d'autres termes, il s'agit de déceler l'importance du rôle que l'indexeur attribue à un élément ou à une combinaison d'éléments par rapport aux autres éléments.

Cette notation n'est pas absolue, mais constitue un moyen d'atteindre une meilleure compréhension des variations interindexeurs.

## 2.5 Hiérarchisation de l'utilisation des éléments de connaissance.

Une façon d'identifier l'importance accordée à un élément de connaissance, en plus du nombre de mentions et d'utilisations de cet élément, consiste à repérer dans les protocoles des énoncés où l'indexeur favorise un élément au détriment d'un autre. Les protocoles ont donc été examinés de nouveau en fonction de l'identification de cette hiérarchie.

Les résultats de cette analyse sont présentés de la façon suivante: un bref rappel de la nature de la prise en compte des divers éléments de connaissance utilisés par l'indexeur, suivi de la hiérarchisation de ces éléments exemplifiée par des énoncés, et finalement, un tableau synthèse (Tableau XX) permettant de comparer l'ordre de priorité accordée aux éléments par chaque indexeur.

### 2.5.1 E1

Globalement, le processus d'indexation de E1 peut être caractérisé par la prise en compte de l'exactitude, des usagers, de la capacité de discrimination, de la rareté et de l'actualité.

Ainsi, pour E1 l'application de l'exactitude se traduit par:

- la prise en compte de l'ampleur du traitement ou du nombre d'occurrences d'une notion;
- la vérification de l'adéquation du sens du terme avec celui du descripteur.

Sa prise en compte de l'utilisateur est reflétée par le choix de termes qui répondent à un grand éventail d'utilisateurs:

- aux besoins d'utilisateurs spécialistes dans le domaine;
- aux besoins d'utilisateurs de la clientèle grand public.

La capacité de discrimination est également un élément important dans son processus d'indexation qui se traduit principalement par :

- le rejet de termes qui ne permettent pas de repérer un ensemble restreint de documents pertinents.
- le choix de termes qui permettent de particulariser le document.

Notons que cette dernière application est étroitement liée à l'actualité, la rareté et l'utilisateur.

L'examen des verbalisations de E1 permet de saisir l'importance qu'elle accorde aux éléments usagers, capacité de discrimination et exactitude. Ces éléments apparaissent comme prédominants lors des décisions de choix ou de rejet de termes. Par ailleurs, l'actualité et la rareté suivent ces premiers dans la hiérarchisation des éléments utilisés par E1. Finalement, l'exhaustivité et la précision se retrouvent au bas de la hiérarchie en tant qu'éléments n'exerçant qu'une influence négligeable sur la prise de décision.

Un premier ensemble d'extraits, relatifs au traitement du concept *modélisation*, rend compte des éléments de connaissance prédominants dans le processus d'indexation de E1, soit la prise en compte de l'utilisateur et la capacité de discrimination, liée ici à l'actualité de la notion. Précisons à propos du traitement de ce concept que E1 évalue les termes *modélisation*, *modèle ISCST*, *logiciel* et le DE *informatique*. Les trois premiers termes ne sont pas dans le thésaurus, ce qui conduit E1 à retenir *informatique* en DE2 et à inscrire *logiciel* et *ISCST* dans le champ des candidats descripteurs.

Par rapport au terme *logiciel*, E1 tient compte de l'évolution du vocabulaire en environnement et souligne la capacité de ce terme à permettre un repérage restreint de notices spécifiques à ce concept pour les usagers intéressés uniquement par cet aspect. Il y a donc prise en compte simultanée de la capacité de discrimination du terme liée aux besoins d'un type particulier d'utilisateurs.

#### **E1 Biogaz**

La seconde évaluation de l'impact sur la qualité a été effectuée en modélisant la dispersion des polluants dans l'atmosphère. *Logiciel* (écrit sur sa feuille de travail). Peut-être que dans le thésaurus on va parler de traité informatique. Il peut y avoir un autre terme pour le *logiciel ISCST*, plus général.

**E1 Biogaz v2**

Parce que de plus en plus on voit des *logiciels* spécialisés dans n'importe quel domaine. Mais en environnement entre autre, pour aider à effectuer telle tâche, telle évaluation. J'imagine que s'il y a des gens qui peuvent retracer cela directement ça leur évite de se taper plusieurs études.

La suite du traitement indique clairement qu'une notion d'actualité est considérée par E1, principalement à cause de sa capacité à discriminer un ensemble restreint de documents au repérage. Une fois de plus, ce choix de terme est étroitement lié à sa prise en compte des requêtes potentielles des usagers.

**E1 Biogaz v2**

Si c'est à la mode de parler des autres effets, tu vas être portée à les mettre. Comme dans vingt ans, *informatique* on ne mettra plus cela, hein! Il va y en avoir partout, ça va être applicable à n'importe quoi, on va laisser faire.

Finalement, le descripteur *informatique* est retenu dans le but de fournir aux usagers un accès à la notice par ce concept. Le champ non interrogeable des candidats descripteurs indiquera à l'utilisateur les termes réellement couverts par le document, soit *logiciel* et *ISCST*.

**E1 Biogaz v2**

*Informatique*, c'est quand même très général. Mais par contre, par expérience, je sais qu'aussitôt qu'il existe un *logiciel*, c'est important de le mettre parce qu'il y en a qui vont faire de plus en plus juste leur recherche sur les *logiciels*. Donc juste en mettant *informatique* ils vont savoir que c'est sûrement un *logiciel* ou une application informatique quelque chose.

L'utilisateur et la capacité de discrimination sont les éléments de connaissance les plus souvent évoqués pour le choix et le rejet des termes évalués par E1, ce qui les place au premier rang de la hiérarchie. Ils se positionnent avant l'exactitude, qui est évoquée régulièrement, mais rarement seule pour l'évaluation d'une notion. D'ailleurs, l'exactitude est parfois mise de côté au profit de l'utilisateur ou de la capacité de discrimination.

Les énoncés suivants démontrent l'utilisation conjointe de l'exactitude et de l'utilisateur pour l'évaluation des termes; le premier énoncé en termes généraux et les suivants en fonction de l'évaluation des différents *sites* pour le document *Étang aéré*, et l'évaluation du descripteur *système d'épuration des eaux usées* pour le document *Compostage*.

**E1 Étang aéré v2**

En lisant, en soulignant ces différents mots significatifs, t'essaies de garder en mémoire celui qui revient le plus souvent, celui qui semble le plus significatif, celui qui serait le plus utilisé par les usagers pour repérer

l'information. Puis par la suite, tu vérifies dans ton thésaurus qu'est-ce qui se rapproche le plus de ce mot-là que t'as choisi de garder.

### **E1 Étang aéré v2**

(*Traitement des noms propres des sites, soit Lufty, Monfette, Boysfarm et Méandre*) Tu vois qu'ils représentent plusieurs pages. Finalement, c'est l'explication du titre. C'est les noms des sites. Ils sont aussi importants que le nom des municipalités. Pour le chercheur tu ne peux pas passer à côté de ça. Comme j'ai dit, ça peut être son seul repère. (...) Ça dépend toujours de la façon que c'est présenté. Si on avait parlé du site *Lufty* juste quatre lignes, j'aurais laissé tomber. D'un autre, on en aurait parlé pendant des pages, c'est sûr que tu le mets.

E1 retient les noms des sites principalement en raison de son souci constant de représenter le document pour le plus grand ensemble d'utilisateurs. La couverture des termes doit néanmoins être suffisante pour justifier le choix.

### **E1 Compostage v1**

Ça me semble être la même chose que boues résiduaires. Qu'est-ce qu'on disait, encore, sur eaux usées? (*Lit dans le thésaurus*) Déchets, eaux industrielles, déchets liquides. Boues d'épuration, et épuration, système d'épuration des eaux usées. C'est pas ça!

### **E1 Compostage v2**

Je sais qu'on avait vu au tout début que tout ça venait du traitement des eaux usées, mais je ne veux pas trop m'égarer. Je ne veux pas amener le chercheur sur un document du traitement des eaux usées, c'est vraiment spécifique aux boues. Un deuxième stade du traitement des eaux usées. Alors c'est pour cela.

L'exemple précédent permet de constater que E1 tient à l'exactitude en relation avec l'adéquation du sens du terme et celui du descripteur, en plus d'évoquer l'utilisateur comme argument supplémentaire pour le rejet du descripteur *système d'épuration des eaux usées*. Notons cependant que E1 revient plus tard sur sa position et inscrit ce descripteur sur sa liste finale puisque aucun autre descripteur ne permet de représenter cette notion essentielle (*boues*) du document.

Par ailleurs, plusieurs indices démontrent la présence de l'utilisateur et de la capacité de discrimination sur l'exactitude, permettant ainsi de rendre compte de certaines différences des listes de E1 en comparaison avec celles des autres indexeurs.

Ainsi, malgré le peu d'exactitude du descripteur *agricole déchet*, E1 le conserve en évoquant l'importance de représenter cette notion pour l'utilisateur. Bien qu'il s'agit ici d'un compromis dû au vocabulaire d'Infoterra, E1 aurait pu rejeter ce terme si l'exactitude avait été l'élément central sur lequel repose son choix.

**E1 Compostage v1**

(Lit dans TH) Agriculture, pour la valorisation agricole... agricole, non! Terre, écologie, économie, gestion, production, bâtiment. Agricole, virgule, déchet. Ce n'est pas clair mais il n'y a pas autre chose. Agricole, déchet.

**E1 Compostage v2**

Mais par contre, il est repris dans *compost et engrais naturel*. C'est le mot *agriculture* qui devrait être là, absolument pour le chercheur qui cherche l'application de l'épandage en agriculture.

De plus, rappelons que plusieurs termes, dont le document traite amplement, tels que *évaluation, échantillonnage, répercussions environnementales, contraintes environnementales, caractérisation, usine, implantation, etc.*, sont rejetés à cause de leur caractère non discriminant pour un document du domaine de l'environnement. E1 soutient également que ce type de termes risque de créer trop de bruit au repérage. Cependant, plusieurs de ces termes se retrouvent sur les listes finales des autres indexeurs.

**E1 Étang aéré v2**

Je les (*les différentes contraintes environnementales*) laisse tomber parce que c'est Envirodoq. Si j'avais eu le même document dans une autre boîte, où la majorité des documents ne sont pas des études environnementales, j'aurais pu écrire comme descripteur, ou DE2, *contraintes environnementales*.

**E1 Biogaz v1**

Campagne *d'échantillonnage*. Ça se retrouve dans la plupart des études environnementales, ce n'est pas vraiment nécessaire de le mentionner.

Par ailleurs, E1 tend à faire ressortir la particularité propre de chaque document, ce qui le rend unique par rapport à l'ensemble des documents de la base de données et permet également à l'utilisateur de repérer rapidement les documents relatifs à cet aspect plus particulier, tels que *Rivière du Nord, ISCST, logiciel, modélisation, informatique, torchère, CUM, revue de littérature, analyse coûts/bénéfices, déchets agricoles, recommandations, rapport de recherche*. Cependant, cette prise en compte de la capacité de discrimination se fait au détriment de l'exactitude puisque plusieurs de ces notions sont peu élaborées dans le document.

Quant aux éléments d'actualité et de rareté, ils ne sont à peu près jamais évoqués seuls et servent d'argument complémentaire à l'utilisateur ou à la capacité de discrimination pour la sélection d'un terme. En conséquence, ils sont classés à la suite des éléments usager, capacité de discrimination et exactitude dans la hiérarchie. Par exemple, rappelons le traitement du terme *informatique* qui a été retenu principalement pour l'utilisateur même si E1 soulève son aspect d'actualité.

Il a été mentionné à plusieurs reprises que E1 ne tient pas à produire une indexation exhaustive. Ce fait est appuyé par la moyenne de descripteurs retenus par E1 (9 DE) pour l'ensemble de ses listes et appuyé par ses verbalisations. Ainsi, l'exhaustivité se retrouve tout au bas de la liste hiérarchique, tout comme la précision d'ailleurs. En fait, nous avons vu aux sections correspondantes que ces deux éléments ne sont que rarement évoqués et lorsqu'ils le sont, leur influence sur le choix de termes est négligeable.

### 2.5.2 E2

De manière générale, la processus d'indexation de E2 peut être caractérisé par la prise en compte des éléments d'exhaustivité associée à la représentation de l'ensemble des facettes, d'exactitude, de l'usager et de l'actualité.

Son application de l'exhaustivité répond aux conditions suivantes:

- la représentation de l'ensemble des facettes ou aspects;
- une limite de six descripteurs par champ.

Le respect de l'exactitude est traduit par deux conditions:

- l'ampleur du traitement de la notion;
- l'adéquation entre le sens du terme et celui du descripteur.

Sa prise en compte de l'usager est reflétée par:

- l'ordonnement des descripteurs pour indiquer à l'usager les relations entre les termes;
- la prise en compte de la satisfaction de l'usager;
- l'inférence du type de mots utilisés par l'usager.

Comme il a été mentionné précédemment, le processus d'indexation de E2 est particularisé par le fait d'indiquer les relations entre les descripteurs en faisant suivre l'un à la suite de l'autre les termes qu'elle veut associer. Il s'agit en fait d'apporter une précision à la façon dont une notion est traitée dans le document qui permettra à l'usager de saisir ces

relations une fois la notice repérée. C'est un reflet de son évaluation du produit global, inséparable de son application de l'exhaustivité.

Lorsqu'on examine l'application des éléments de connaissance utilisés par E2, il ressort de manière suffisamment claire que certains éléments sont évoqués plus souvent pour l'évaluation de la pertinence d'un terme. Ainsi, l'exactitude et l'exhaustivité apparaissent comme les éléments premiers dans la hiérarchie de E2.

Les énoncés suivants servent à illustrer ce constat. Rappelons que cette hiérarchie ne se veut pas absolue, mais sert à démontrer qu'une pondération différente dans l'utilisation des divers éléments de connaissance entraîne des décisions différentes.

Pour le document *Biogaz*, E2 tient à représenter un des aspects du *biogaz* présenté dans le document, soit la *caractérisation* du biogaz. Ce terme n'étant pas dans le thésaurus, E2 tente de trouver un synonyme, le descripteur *identification* est alors considéré.

#### **E2 Biogaz v2**

Je cherchais un mot synonyme parce que je savais que je ne pouvais pas prendre *caractérisation*, que j'ai mis en candidat. Et *identification* me semblait synonyme...

*Identification* c'est plus le nom qu'on donne à trouver quel genre de gaz, identifier. *Caractérisation* c'est plus que cela, c'est comment se comporte ce gaz, c'est plus le caractère propre au gaz. Donc ce n'est pas la même chose du tout. Ça ne me satisfait pas...

Quand ça ne me satisfait pas, j'aime quasiment mieux rien mettre que mettre un terme qui est mauvais, qui ne veut pas dire la même chose.

Cet exemple démontre la priorité du respect de l'exactitude sur l'exhaustivité. L'adéquation du sens des mots ne correspond pas, ce qui entraîne le rejet du descripteur évalué. Notons que le terme *caractérisation* est conservé en candidat descripteur, de ce fait, l'exactitude et l'exhaustivité se trouvent toutes deux appliquées en égard à cette notion.

L'épisode suivant constitue un autre exemple de la priorité de l'exactitude sur l'exhaustivité, mais cette fois, reliée à l'ampleur du traitement de la notion évaluée.

#### **E2 Énergie v2**

expé

Pourquoi *équipement* t'as frappé?

E2

Parce que s'il y avait eu une partie assez importante sur l'*équipement* que ça prend pour produire tout cela (*les différentes énergies renouvelables*). C'est un essai probablement qu'ils ont fait, je l'aurais probablement

pris.... Parce qu'ils disent, (*lecture dans le résumé*) dans un premier temps, un deuxième temps et une présentation de l'équipement proposé ainsi qu'un scénario possible suivant ces deux premières parties pour résoudre les problèmes énergétiques. Je pensais que c'était comme une troisième partie (*donc aussi importante que les deux premières*), mais cette partie-là, je me suis aperçue que ce n'était vraiment pas développé.

Ainsi, E2 décide de ne pas représenter la dimension *équipement* du document *Énergie* à partir de la vérification, dans la table des matières, de la couverture que le document fait de cette notion. En effet, la dimension *équipement* n'occupe que deux pages du document. Ainsi, même si cette dimension apparaissait intéressante, E2 a privilégié l'exactitude plutôt que l'exhaustivité.

L'épisode suivant concernant le traitement de la notion *économie* du document *Compostage* démontre la priorité accordée à l'exhaustivité au détriment de l'exactitude qui se rapporte à l'adéquation du sens entre terme et descripteur. Le terme est traité amplement dans le document et constitue un aspect de l'exploitation du compostage que E2 tient à représenter. Les descripteurs relatifs à cette notion ne rejoignent pas tout à fait le sens de la notion du document, mais E2 décide de représenter cette facette par le descripteur *développement économique*. D'ailleurs, le dernier énoncé laisse clairement sous-entendre que E2 a fait un compromis d'exactitude pour pouvoir représenter cette dimension du document.

#### **E2 Compostage v1**

(*Lecture dans le document*) Planification économique, zonage. L'aspect économique, l'aspect commercial. Les débouchés pour le compost.

#### **E2 Compostage v2**

**expé**

Pourquoi tu le relis?

**E2**

Pour plus chercher à spécifier *l'économie*. *L'économie* de quoi peut-être. L'aspect économique était assez bien développé, je voyais tout ce qui parlait de l'aspect économique.

#### **E2 Compostage v2**

Vu que les deux mots étaient associés (*économique, développement*), le mot développement m'embêtait, mais je pense que je n'avais pas le choix. Donc je suis allé voir (*dans le document*) si c'était une forme de développement et quand j'ai vu établissement du projet, projet technique et que je suis allé vérifier à l'intérieur et que c'était vraiment un projet, ça justifiait plus le choix.

De toute évidence, ce compromis est lié au contenu du thésaurus Infoterra, mais si l'exactitude primait sur tout autre élément, E2 aurait pu tout simplement abandonner la représentation de cette notion plutôt que de mettre un terme inexact. Nous verrons

d'ailleurs comment, dans une situation similaire, E3 rejette certaines notions qui ne peuvent être représentées de manière exacte et précise.

Un dernier exemple tiré de l'indexation du document *Biogaz* permet de constater la priorité de l'exhaustivité sur l'exactitude. La notion centrale du document, *biogaz*, n'est pas dans le thésaurus. E2 tente donc de compenser par un terme qu'elle sait ne pas être équivalent. Il s'agit d'une entorse à l'exactitude dans le but de représenter une notion primordiale.

#### **E2 Biogaz v2**

Je garde *gaz carbonique* quand même. Parce que pour moi *biogaz*, c'était des *gaz carboniques*, une forme de *gaz carbonique* mais sous une forme générale pas spécifique.

#### **E2 Biogaz v2**

*Biogaz*, c'était comme une forme de *pollution de l'air*. Mais quelqu'un qui cherche *pollution air* pouvait ressortir ce genre de forme-là. C'est plus général mais c'est dans ce but-là que je l'ai sorti quand même.

#### **E2 Biogaz v2**

*Pollution air* je l'ai marqué parce que je regarde le titre. Le fait de ne pas pouvoir mettre *biogaz*, mettre *pollution air* c'était comme pour me contenter. J'avais besoin de ces mots-là pour compenser. Il y a un lien entre les deux (*biogaz et pollution air*) mais ce n'est pas du tout la même chose.

#### **E2 Biogaz v2**

J'ai trouvé utile de le mettre en DE1 (*pollution air*). Au moins *pollution air* et *gaz carbonique*, ces deux-là me permettaient *élimination du biogaz*. Quand j'ai vu que je n'étais pas satisfaite de mes mots, je vais le garder en DE1. Pour essayer de faire ressortir le fait de *l'élimination du biogaz* que je ne pouvais absolument pas mettre.

Ainsi, E2 conserve sur sa liste finale deux descripteurs reliés à la représentation de la notion de *biogaz*, soit *gaz carbonique* et *pollution air*.

Ce que l'on peut conclure de l'analyse de ces extraits est que E2 mise principalement sur les critères d'exhaustivité et d'exactitude, mais que l'exhaustivité prévaut lorsque le vocabulaire du thésaurus ne permet pas une représentation exacte d'une dimension qui se rapporte au sujet premier du document. En effet, le descripteur *identification* est un terme secondaire qui vient qualifier le *biogaz*, la notion d'*équipement* vient préciser les équipements proposés pour résoudre les problèmes en énergie de cette région, mais ne constitue pas le sujet central. Au contraire, la dimension *économique* du projet de compostage occupe un chapitre de l'étude et la *pollution de l'air* constitue l'impact majeur sur l'environnement causé par l'élimination du *biogaz*. Notons cependant que E2 retient

*pollution de l'air* presque comme substitut au terme *biogaz* qu'elle ne peut représenter par aucun descripteur satisfaisant, lequel terme est le sujet principal du document.

Ainsi, pour la pondération des divers éléments de connaissance utilisés par E2, on peut conclure que l'exactitude et l'exhaustivité occupent conjointement le premier rang. Précisons que son souci de représenter les différentes facettes est considéré au même niveau d'importance puisqu'il s'agit d'une application directe de l'exhaustivité permettant de caractériser E2.

À travers les extraits de protocoles précédents, nous pouvons également remarquer que le critère de précision n'est évoqué implicitement qu'en relation avec l'exactitude, mais n'est pas pris en compte pour le choix des descripteurs. En effet, bien que certains descripteurs évalués ne soient pas du même niveau de spécificité que la notion à représenter, E2 n'évoque pas cette contrainte et argumente plutôt en fonction de l'exactitude ou de l'exhaustivité. À ce propos, revenons brièvement sur le traitement du concept d'*économie*. E2 accepte de prendre le descripteur *développement économique* malgré le fait que le sens du terme d'origine (*aspect économique*) est trahi. En fait, lorsque E2 a déterminé qu'un aspect est important et qu'il doit être représenté dans la liste finale de descripteurs, elle tentera de sélectionner le descripteur le plus adéquat, qu'il exprime ou non de manière précise le terme initial. Le choix des descripteurs *gaz carbonique* et *pollution air* constitue également un exemple de l'influence négligeable de la précision dans le processus global. De ce fait, la précision doit être placée au bas de la hiérarchie des éléments de connaissance utilisés par E2.

L'usager occupe également une place de choix dans la hiérarchie des éléments de connaissance utilisés par E2. Toutefois, l'usager ne semble pas être pris en compte de manière isolée mais plutôt conjointement à d'autres éléments de connaissance, dont principalement les critères privilégiés d'exactitude et d'exhaustivité.

Un premier extrait illustre la combinaison des éléments exactitude et usager. Dans la première partie de l'extrait, E2 souligne que le document ne traite pas expressément de l'aspect *agriculture*, mais mentionne plus loin l'importance de cette dimension. En fait, il semble que E2 oscille entre le respect de l'exactitude – conduisant au rejet de ce terme – ou

celui de l'exhaustivité, en représentant un aspect supplémentaire du document. En tenant compte de la satisfaction de l'utilisateur, E2 opte alors pour le rejet de cette notion, malgré le fait qu'elle lui semble intéressante pour la représentation globale du sujet du document.

#### **E2 Compostage v2**

*Agriculture*, ça me pesait. Il ne parle pas vraiment d'*agriculture* mais le but de *compostage* et ces *boues* c'est en fonction de *l'agriculture*. Je me demandais si j'allais le mettre quand même.

#### **E2 Compostage v2**

Je retourne à la table des matières ou d'autres pages qui pourraient m'aider. Pour vérifier de nouveau si je devais mettre ce mot (*agriculture*). C'est difficile ce genre de situation. C'est dans le but de, mais si on met *agriculture* et un utilisateur veut avoir des documents sur *l'agriculture* et on lui sort un document sur la *gestion des boues résiduaires urbaines...*

C'est pour cela que j'hésitais.

Mais, en même temps, il a quand même une importance à *l'agriculture*. On est pris entre deux idées....

Si je ne l'ai pas mis c'est parce que je me suis mise à la place de l'utilisateur qui allait trouver ça.

#### **E2 Étang aéré v2**

Des fois juste un paragraphe sur un sujet ça frustre les utilisateurs. Dans ce temps-là, on néglige de le mettre.

Nous avons souligné précédemment qu'un aspect de l'exhaustivité considéré par E2 consiste à limiter le nombre de descripteurs à six descripteurs pour chaque champ. Cette limite arbitraire est en fait motivée par le souci de faciliter la lecture de la notice à l'utilisateur.

En ce sens, l'évaluation de la représentation de certaines facettes d'un document se fait conjointement à la prise en compte de l'utilisateur. Il arrive ainsi que les notions soient retenues ou rejetées selon que E2 considère que le terme aidera ou non l'utilisateur à mieux saisir le contenu du document, en évaluant la satisfaction de l'utilisateur au repérage, ou en inférant le type de termes utilisés par l'utilisateur pour le repérage.

#### **E2 Biogaz v2**

C'est surtout le fait de *bilan*. Je pense qu'il existe tout seul comme descripteur, mais ce n'est pas vraiment des mots que les gens vont taper pour essayer de trouver quelque chose. Je ne trouve pas ça utile des mots comme cela. À moins que le titre s'appelle *bilan d'énergie*, mais pas dans ce document-ci où il n'y a qu'une petite partie consacrée.

Notons au passage que le terme *bilan* couvre le même nombre de pages ou plus que d'autres notions retenues en fonction de l'ampleur de la couverture, par exemple *échantillonnage*, *caractérisation* et *pollution air*. On voit bien alors que la prise en compte de l'utilisateur influence la décision, ou du moins est mentionné comme argument additionnel pour l'évaluation d'un terme.

**E2 Biogaz v2**

*Historique des campagnes (qui fait l'objet d'une section de la table des matières). Non, je ne pense pas qu'un usager veuille savoir l'historique, comment ça s'est produit, comment ça s'est fait, qu'est-ce qui est arrivé?*

**E2 Étang aéré v2**

*(E2 considère les noms des différents sites évalués, chacun constituant une sous-section du même chapitre) Lufty, Méandre. Ça ne veut rien dire cela. J'avais vu que ça ne servait à rien de nommer ça. L'information qu'il y avait là, je ne la trouvais pas utile.... Ça ne sert à rien à l'utilisateur de savoir qu'un site s'appelle Boysfarm.*

**E2 Étang aéré v2**

**expé**

Pourquoi tu mets *traitement eau usée*?

**E2**

Parce que s'il y a des usagers qui demandent des documents sur *traitement eau usée*, mais ils ne savent pas que les *étangs aérés* c'est une technologie pour ça (*pour le traitement des eaux usées*). Là on va sortir le document par *traitement eau usée* qu'on va avoir mis, ça va donner encore plus de possibilités, d'avoir plus de documents ou des technologies, ça va apporter plus aux usagers. Et je trouvais aussi que *étang aéré* est peut-être trop spécifique.

**E2 Biogaz v2**

*(Échantillonnage) En DE2 parce que c'est une partie du document. En DE1 si c'est une analyse en laboratoire et que le document possède une grosse partie sur les résultats qu'ils ont fait en échantillonnage.... Mais ce n'est pas le but du volume que j'avais. Je l'ai mis quand même en voulant dire à l'utilisateur qu'il y a des résultats d'échantillonnage à l'intérieur mais ce n'est pas au premier niveau.*

Ce dernier exemple présente une autre dimension de la prise en compte de l'utilisateur par E2. Il s'agit de l'ordre de présentation des descripteurs. En effet, E2 tente de placer les termes en fonction des relations existantes entre les notions, ce qui a pour but de fournir une description plus précise du sujet à l'utilisateur, une fois qu'il aura repéré la notice.

Finalement, précisons qu'à l'étape de révision de son indexation, E2 évoque régulièrement la satisfaction de l'utilisateur comme critère d'évaluation de la pertinence des descripteurs.

**E2 Étang aéré v2**

À ce moment-là (*lors de la révision*), je me mets vraiment à la place de l'utilisateur encore plus, parce que je me dis, c'est comme si ce volume-là était déjà dans la banque et quelqu'un voudrait le rechercher. Je relis en fonction de l'utilisateur, s'il serait satisfait d'après les termes qu'il y a là. Si le document correspond aux termes, et tout ça.

Ainsi, dans la hiérarchie des éléments de connaissance utilisés par E2, l'utilisateur se trouve tout près de l'exactitude et de l'exhaustivité, mais n'est pas considéré d'égale importance puisqu'il n'est pas évoqué comme élément premier pour justifier le choix ou le rejet d'un terme. Il semble que le rôle de la prise en compte de l'utilisateur consiste à amplifier, à ajouter du poids dans l'argumentation en faveur du choix ou du rejet d'une notion évaluée.

La rareté et l'actualité ne sont mentionnées que de manière anecdotique dans les protocoles de E2 et conséquemment doivent être placées au bas de la hiérarchie.

### 2.5.3 E3

Le processus d'indexation de E3 peut être caractérisé par la prise en compte de l'exactitude, de la précision et de la rareté d'une notion et par la représentation du propos central du document. De plus, E3 mentionne régulièrement l'utilisateur.

Pour E3, l'application de l'exactitude est traduite par:

- la prise en compte de l'ampleur du traitement du terme,
- la vérification de l'adéquation du sens du terme avec le sens du descripteur.

Pour produire une indexation précise :

- E3 évalue la correspondance du niveau hiérarchique des descripteurs évalués avec les termes à représenter.

Ses mentions de l'utilisateur se rapportent à:

- l'évaluation de la satisfaction de l'utilisateur au repérage, principalement à l'étape de révision;
- l'évaluation des termes en fonction des termes potentiellement utilisés par les utilisateurs dans leurs requêtes.

E3 semble également s'attarder à la rareté d'un terme en fonction du fonds documentaire du MENVIQ.

Le processus de E3 est également particularisé par l'importance qu'elle accorde aux termes exprimant le propos central du document, évoqué généralement par l'expression "but du document".

La hiérarchie des éléments de connaissance utilisés par E3 est relativement claire. La préoccupation principale de E3 est de représenter les notions qui se rapportent au propos central du document, ce qui limite l'impact que peuvent avoir les autres éléments sur le choix des termes. Ainsi, bien que l'exactitude soit un critère privilégié par E3, elle est évaluée parallèlement au respect de la représentation du propos central du document, ce qui entraîne le rejet de certaines notions exactes, mais non représentatives du propos central. En ce sens, l'exactitude arrive deuxième dans la hiérarchie des éléments utilisés par E3.

Les énoncés suivants illustrent le fait que E3 reconnaît les aspects secondaires, notés et retenus par plusieurs indexeurs, mais les rejette puisqu'ils ne correspondent pas au propos premier.

#### **E3 Étang aéré v2**

Si on fait l'étude de sites d'implantation, le fait que les eaux usées se jettent dans la *Rivière du Nord* ç'a plus ou moins d'importance. Que les eaux usées se déversent dans la *Rivière du Nord* ou dans une autre, ce n'est pas le but, ce n'est pas le document. Je me suis dit que ce n'était pas le document.

#### **E3 Biogaz v2**

*Échantillonnage* ç'a rapport à la caractérisation. Quand tu veux caractériser, t'échantillonne pour savoir de quoi est composé.... On peut faire de l'échantillonnage mais c'est secondaire, c'est dans le but de nous permettre d'évaluer les impacts. Je l'ai laissé tomber. Le document n'est pas sur *l'échantillonnage*.

#### **E3 Énergie v1**

(*Lecture dans la table des matières*) Aperçu sur les milieux physique et humain.

#### **E3 Énergie v2**

Ce n'est pas vraiment le but du document. Ils donnent les conditions générales de la place. Donc si on cherche avec un DE géographique, on va quand même repérer ces données-là.

Par ailleurs, la prédominance de la représentation du propos central et de l'exactitude dans le processus d'indexation de E3 fournit l'explication à la non exhaustivité de ses listes.

La précision est placée à la suite de ces deux premiers éléments privilégiés. En effet, on remarque que la précision est régulièrement prise en compte dans l'évaluation des notions traitées. Cependant, un terme doit répondre d'abord aux conditions d'exactitude pour que E3 recherche le descripteur le plus précis pour le représenter. En ce sens, la précision est dépendante de l'exactitude puisque E3 peut connaître le descripteur précis permettant de représenter une notion du document, mais cette notion est généralement

rejetée si elle n'est pas jugée exacte. Soulignons que la précision demeure toutefois une condition pour le choix final des descripteurs.

L'énoncé suivant illustre le rejet du descripteur *station d'épuration*, malgré le fait qu'il s'agisse du descripteur à employer pour (EP) représenter le terme *usine d'épuration*. L'exactitude et la représentation du propos central sont toutes deux évoquées pour le rejet du descripteur.

### **E3 Étang aéré v2**

**expé** Pourquoi t'as enlevé *station d'épuration*?

**E3** Ici on disait, la grille d'évaluation des sites fait voir qu'aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs. Donc là vraiment, le document il portait sur l'implantation d'étangs, il ne portait pas sur la construction d'une *usine*. C'est pour cela que je l'ai enlevé. Je me disais qu'*usine* on en parle par-ci par-là, mais ce n'est pas le document, ce n'est pas vraiment ça.

En termes plus généraux, l'énoncé suivant fournit une explication au comportement global de E3 en fonction du propos central et de l'exactitude.

### **E3 Étang aéré v2**

Parce que des fois, le document nous fait penser à quelque chose, nous fait penser à un sujet, indirectement cela a rapport à ce sujet-là. Sauf qu'on n'en parle pas vraiment dans le document. Je trouve dans ce temps-là qu'on n'est peut-être pas justifié de le mettre. Parce que si une personne cherche quelque chose avec le sujet et il retrouve le document et on n'en parle pas vraiment. Mais ça a un lien quelconque avec. Il ne sera peut-être pas content de l'avoir. Parce que ça fait du bruit.

L'énoncé suivant permet de souligner le souci de précision et l'influence de ce critère sur le choix des descripteurs finaux. Notons que le terme traité (*biomasse*) est déjà considéré exact et représentatif du propos.

### **E3 Énergie v2**

(Lors de la consultation du thésaurus, E3 évalue la précision des descripteurs *énergie de la biomasse et bouse comme combustible* pour représenter le terme *biomasse*).

**expé** Même si *énergie de la biomasse* était dans le titre, tu préférerais *bouse comme combustible*?

**E3** Oui, c'était plus précis. Mais après je suis retourné vérifier dans le document et c'est vraiment des *bouses*, humaines ou animales. *Biomasse* c'est plus vague. (Dans le thésaurus) *Biomasse, terme spécifique (TS) déchet employé comme combustible*, je trouvais que c'était trop général. J'ai pris *bouse* parce que là-dedans (dans le document) on parle vraiment de *bouse*.

Dans l'énoncé suivant, l'évaluation de la satisfaction de l'utilisateur prise en compte par E3 lors de la révision de sa liste de descripteurs renvoie de manière implicite à la précision de son indexation, c'est-à-dire à la capacité des descripteurs choisis à réduire aux seuls documents très pertinents les documents repérés.

**E3 Énergie v1**

Descripteurs secondaires, **sources d'énergie non-polluantes**, dans les pays en voie de développement. Si quelqu'un cherche avec ça, il va être intéressé à voir ce document.

**E3 Énergie v2**

Je me dis, si une personne cherche avec ça, va-t-elle être contente d'avoir ce document-là? Si la réponse est oui, bien, pas pire. Je peux en oublier, mais au moins ceux qui sont là, ils risquent d'être pertinents.

Étant donné la priorité accordée au propos central du document et à l'exactitude, les listes de termes de E3 répondent ainsi à un aspect de la précision. Mais puisque cette précision découle directement de l'appréhension explicite des éléments de connaissance prioritaires, la précision doit être placée après l'exactitude dans la hiérarchie.

Rappelons que la prise en compte de la précision et de l'exactitude apparaît de manière plus marquée lors de la consultation du thésaurus Infoterra, dont le vocabulaire est inconnu des indexeurs. En effet, E3 consulte en profondeur les différentes sections de ce thésaurus afin de sélectionner le descripteur le plus approprié.

L'utilisateur est pris en compte, mais ne joue pas un rôle prépondérant pour le choix des termes. En fait, l'utilisateur est rarement évoqué seul et sert principalement d'argument supplémentaire pour la sélection ou le rejet d'une notion déjà jugée primordiale ou secondaire, c'est-à-dire dont le choix ou le rejet a déjà été décidé.

**E3 Étang aéré v2**

(E3 évalue la pertinence des termes *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*) Si quelqu'un cherche des documents sur *St-Sauveur-des-Monts*, et qu'il tombe là-dessus, peut-être qu'il ne sera pas nécessairement content d'avoir ça. Parce qu'on n'en parle pas tant que ça de *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*.

En plus de ne pas avoir un impact réel sur le choix des termes, il semble que l'utilisateur devra adapter sa stratégie de recherche à celle que lui prête E3. L'énoncé suivant indique clairement ce fait, tout en soulignant la prédominance de l'exactitude et de la représentation du propos sur l'utilisateur pour le choix de termes.

**E3 Biogaz v1**

Caractérisation du biogaz. *Caractérisation.* Devrais-je le mettre?

**E3 Biogaz v2**

*Caractérisation* occupe une bonne partie du document, sauf que le but de ce document c'est une étude d'impact. J'ai résolu en me disant ça mais pas nécessairement de caractériser le *biogaz*. Non, je le mets pas. Si quelqu'un cherche la *caractérisation* du *biogaz* et ne trouve rien, il va aller voir à *biogaz* et va sortir ça et donc va l'avoir quand même.

La précision prime également sur la prise en compte de l'utilisateur et ce dernier devra une fois de plus adapter sa stratégie de recherche à la façon d'indexer de E3.

### **E3 Énergie v2**

expé

Pourquoi t'hésites tant que ça avec *énergie de la biomasse*?

E3

À cause du titre. Parce qu'ici (*dans le titre*) on parle de *l'énergie de la biomasse*. Mais finalement, je suis allé au plus précis. J'aime mieux que ce soit précis.... En me disant que là-dedans (*thesaurus*), si l'utilisateur va voir à *énergie de la biomasse*, il va voir aussi *boise comme combustible (en terme associé)*, il peut toujours l'essayer. De toute façon, il est là-dedans, le renvoi est là.

Finalement, le dernier énoncé démontre que l'utilisateur n'est pas réellement pris en compte dans le processus d'indexation de E3, même s'il est mentionné. En effet, E3 conserve un terme jugé important par l'ampleur du traitement, malgré qu'elle suppose qu'un utilisateur n'utiliserait pas ce terme dans sa requête.

### **E3 Étang aéré v2**

Ici on voit la grosse partie qui parle de *l'évaluation des sites*. Donc si quelqu'un fait une recherche vraiment poussée et il veut avoir *l'évaluation* d'emplacement pour telle construction, peut-être qu'il va le mettre.... Mais je ne suis pas sûre qu'une personne va chercher aussi spécifique que ça.

Étant donné le peu de poids que semble accorder E3 à l'élément utilisateur, il semble adéquat de placer cet élément parmi les éléments de connaissance n'exerçant que peu ou pas d'influence sur le choix de termes.

La rareté d'un terme associée au fonds documentaire du centre est également un élément pris en considération. Cependant, l'importance de cet élément sur le processus global d'indexation est relative puisque E3 ne l'utilise qu'une fois en choisissant le descripteur *bruit* pour le document *Biogaz*. Ainsi, nous plaçons la rareté tout au bas de la hiérarchie des éléments pris en compte par E3, même si nous inférons que si E3 avait rencontré d'autres termes de ce même statut, elle les aurait sans doute traités.

## **2.5.4 E4**

L'ensemble des éléments de connaissance utilisés par E4 est relativement restreint. Ainsi, le processus d'indexation de E4 peut être caractérisé par une prise en compte marquée de l'exhaustivité et par la considération des notions d'actualité et des notions plus

rarement rencontrées dans le domaine de l'environnement. L'exactitude figure également parmi les critères mentionnés.

L'application de l'exhaustivité se traduit par:

- l'extraction de toutes les notions (ou termes) provenant des parties les plus informatives du document, ainsi que des notions mises en relief par une typographie particulière;
- la recherche de descripteurs permettant de représenter chacun des termes retenus à la première étape de l'indexation.

E4 s'attarde aux termes nouveaux du domaine ainsi qu'aux termes rarement rencontrés. Une fois repérés, ils sont retenus en brouillon et généralement retenus en candidats descripteurs s'ils n'ont pas d'équivalents dans le thésaurus.

Quant aux autres éléments de connaissance, il a été souligné précédemment que E4 ne leur accorde pas une priorité dans son processus.

Parmi les éléments de connaissance utilisés par E4, l'exhaustivité est sans contredit en tête de liste. Les allusions faites à propos de la précision, de l'exactitude et de l'usager ne prévalent pas sur l'exhaustivité. E4 retient donc un grand nombre de termes comme première représentation et tentera de traduire tous ces termes lors de la consultation du thésaurus, peu importe leur réelle représentativité du contenu du document.

Ainsi, les énoncés suivants mettent de l'avant la prédominance de l'exhaustivité sur la précision et l'exactitude. Par exemple, E4 représente les termes *composition* et *pluie* par des descripteurs de nature plus générique, soit *analyse chimique* et *précipitations*.

**E4 Biogaz v1**

(Consultation du thésaurus) **Analyse chimique.** Voilà! OK. Cela fait pour *composition*.

**E4 Biogaz v2**

C'est un mot qui vient de déclencher. T'analyses la *composition*. Quelque chose qui commencerait par *analyse* et qui aurait rapport avec la composition de mes gaz.

**E4 Biogaz v2**

*Composition* n'est pas au même niveau. Mais *analyse* englobe *composition*.

**E4 Énergie v1**

(Consultation du *thésaurus*) Atmosphère, je n'ai pas besoin de ça. Phénomènes atmosphériques, ah! Peut-être? Précipitations. Précipitations. Voilà! OK! (*raye pluie des termes brouillon, précipitations mis pour pluie*)

En fait, un descripteur peut être sélectionné dès qu'il possède un lien, aussi ténu soit-il, avec la notion retenue en brouillon, résultant souvent en une représentation non exacte et non précise. L'extrait suivant illustre le fait que E4 ne remet pas en question l'adéquation du sens du terme et du descripteur pour juger de la valeur d'un descripteur.

**E4 Énergie v1**

(Consultation du *thésaurus*) Pollution, là on ne parle pas de ça. Climat et changement climatique. Changement climatique, ça, je le mets.

**E4 Énergie v2**

expé

À quoi tu fais référence quand tu mets cela?

E4

Parce qu'ils parlaient dans le résumé qu'il y avait un problème de *désertification*, accentué par la *monoculture*, quelque chose comme cela.

L'importance d'une représentation exhaustive au détriment de la précision et de l'exactitude est mise en évidence par les extraits suivants. Ce n'est pas le fait d'avoir ou non représenté l'ensemble du document qui pousse E4 à ajouter ou chercher d'autres DE, mais bien le fait qu'elle trouve qu'elle n'en a pas suffisamment. En effet, E4 décide d'ajouter des descripteurs lorsqu'elle constate que sa liste ne comporte que six DE1. Il apparaît ici que le souci d'exhaustivité de E4 est lié avant tout au nombre de descripteurs plutôt qu'à la représentation des divers aspects du document.

**E4 Compostage v1**

(Réviser sa liste de descripteurs) Il y a récupération des déchets, composts et engrais naturels, amélioration du sol, recyclage. Là, je me rends compte que je les ai tous mis en descripteurs principaux.

**E4 Compostage v2**

J'étais tellement contente d'en avoir trouvé, pas beaucoup, six, que je les ai tous placés en descripteurs principaux.

**E4 Compostage v1**

(Relit sa liste de descripteurs) J'en ai aucun secondaire (DE2)! Ça ne se peut pas! Je retourne à la liste des termes d'après leur numéro de code (*dans le thésaurus*). Bon. Atmosphère, non! Lithosphère, non! Sol. Terres agricoles. Terres alcalines, terres contaminées, terres de bruyère... Alors, je pourrais mettre ici, en secondaire (DE2) terres agricoles.

**E4 Compostage v2**

expé

Pourquoi *terre agricole* en secondaire?

E4

C'est carrément que je n'avais absolument rien dans les descripteurs principaux qui touche à la *valorisation agricole* que j'avais en tête au départ. Je viens de spécifier un peu plus que c'est pour les *terres agricoles* tous ces procédés que j'ai mis en descripteurs principaux.

Par ailleurs, l'application particulière de l'exhaustivité de E4 entraîne la sélection de descripteurs qui ne se rapportent pas au sujet traité par le document, mais qui décrivent en détail le contexte relatif à l'étude. Par exemple, les notions *ensoleillement*, *désertification* et *capacité optimales du sol* ne sont pas traitées dans le document *Énergie*, mais rendent compte du contexte unique de ce pays:

**E4 Énergie v1**

(Lecture du résumé) ... les sols dior. sur lesquels poussent facilement l'arachide. L'ensoleillement du site y est très important, ce qui accroît le phénomène de désertification. En tout cas, **ensoleillement, désertification.**

**E4 Énergie v2**

**expé**

Pourquoi ces mots-là?

**E4**

C'est juste par rapport au caractère physique de ce coin de pays là qui est situé au Sahel. Ce sont tous des phénomènes qui sont importants là-bas.

**E4 Énergie v1**

(Consultation du thésaurus) Écosystèmes terrestres, sols, terres agricoles, alcalines, contaminées, de bruyère, capacités optimales du sol. Ouais, peut-être ça? **capacités optimales du sol.**

**E4 Énergie v2**

Ça je le garde parce que dans ce coin-là, c'est important de tirer du sol le plus qu'il peut te donner.

En résumé, l'exhaustivité pratiquée par E4 ne répond à aucune contrainte d'exactitude, de précision, ce qui explique le nombre élevé de descripteurs.

L'exhaustivité est l'élément de connaissance qui domine le processus d'indexation de E4. Cependant, on retrouve une considération secondaire mais présente pour les éléments d'actualité, de rareté et de nouveauté. E4 semble en effet accorder une attention spécifique à ces aspects qu'elle placera dans le champ des candidats descripteurs si aucun descripteur officiel ne permet de rendre le terme du document. En conséquence, nous plaçons les deux éléments de rareté et d'actualité à la suite de l'exhaustivité dans la hiérarchisation des éléments de E4. Quant à l'élément de nouveauté, il doit venir à la suite de l'actualité et de la rareté puisqu'il ne rend compte que du choix d'un seul terme (*design*).

**E4 Biogaz**

Le biogaz constitue une nuisance environnementale. Le méthane qu'il contient est inflammable et reconnu pour favoriser l'effet de serre. **L'effet de serre.**

**E4 Biogaz v2**

C'est une chose dont on entend beaucoup parler, il y a beaucoup d'écrits là-dessus. Tu ne passes pas à côté d'un document qui parle *d'effet de serre* sans le souligner.

**expé**

Mais est-ce que le document en parle *d'effet de serre*?

**E4**

Là je ne sais pas à ce niveau-là.

**E4 Biogaz v1**

Une partie du biogaz servira à alimenter une usine de cogénération d'électricité. Cogénération.

**E4 Biogaz v2**

*Cogénération* c'est dans l'air, on en parle beaucoup.

expé

Qu'est-ce que ça veut dire?

**E4**

Je ne le sais pas. Hydro-Québec en parle beaucoup, ils veulent faire des usines de *cogénération*. Peut-être un déchet qui est valorisé, utilisé avec autre chose pour produire de l'électricité, peut-être. C'est dans l'air.

**E4 Étang aéré v2**

C'est la première fois que j'entends parler de *design d'étang*. C'est pour cela que je l'ai laissé, parce qu'il m'a intrigué, mais en même temps je me demandais si ce n'était pas plus ou moins secondaire et même moins que secondaire.

Ces énoncés illustrent bien que la prise en compte des éléments d'actualité et de rareté précèdent les critères d'exactitude et de précision dans la hiérarchie et contribuent à l'élaboration d'une représentation exhaustive. E4 retient *effet de serre* et *cogénération* malgré le fait que le document ne traite pas spécifiquement ou abondamment de ces notions. En fait, E4 ne vérifie pas l'ampleur du traitement de ces termes avant de les sélectionner. De même, le terme *cogénération* apparaît dans une seule phrase de l'introduction, mais se retrouve parmi la liste des candidats descripteurs.

De plus, on remarque que les termes d'actualité et des termes inusités ont préséance sur d'autres termes de même niveau qui ne sont pas retenus en brouillon ou qui sont abandonnés à l'étape de traduction car non présents dans le thésaurus. À l'opposé, les notions rares et d'actualité seront retenues en candidats descripteurs. Ceci nous permet de placer les éléments d'actualité et de rareté au second rang des éléments de connaissance utilisés par E4.

Les énoncés suivants démontrent que des termes de même niveau sont traités différemment selon qu'ils représentent ou non une notion d'actualité.

**E4 Biogaz v1**

Le biogaz est surtout composé de méthane et de bioxyde de carbone. Il contient également en trace plusieurs composés organiques volatils. OK.

**E4 Biogaz v2**

*Méthane*, je me rappelle que c'était un descripteur très en vogue, c'est-à-dire un candidat, parce qu'il n'était pas dans le thésaurus. Il y avait beaucoup d'études sur les effets du *méthane*. *Bioxyde de carbone*, ça ne me disait rien donc je ne l'ai pas mis.

Par ailleurs, l'énoncé suivant permet de saisir un peu plus la démarche suivie par E4.

**E4 Étang aéré v2**

Je cherche comment le rendre (*les concepts de contraintes environnementales et physiques*) et les mots qui sont là ne m'inspirent pas. La page précédente, chacun des mots me saute au visage. Mais là, pas du tout.

On peut donc supposer que son application de l'exhaustivité suit un processus relativement aléatoire qui est, d'une part tributaire de la structure et de la présentation de l'information dans le document, et d'autre part, qui repose également sur sa connaissance du domaine et de son évolution. En effet, E4 semble s'attarder à toute notion qui "l'accroche", soit parce qu'il s'agit d'un terme inusité, d'actualité, soit parce que le terme lui est familier ou parce qu'elle sait que le terme est également un descripteur, ou parce qu'un terme est mis en évidence par la mise en forme du document. Par ailleurs, on remarque que E4 retient certains candidats descripteurs (*expropriation* et *zone inondable* pour le document *Étang aéré*) parce qu'elle sait que le thésaurus ne couvre pas suffisamment cette dimension.

Cette appréhension unique de l'exhaustivité déborde amplement de la prescription de la norme en ce qui concerne l'application de ce critère. Non seulement E4 n'évalue pas la valeur potentielle des termes pour les usagers et ne tente pas de cerner les notions réellement traitées dans le document, mais elle retient un grand nombre de termes qui ne permettent pas de réduire le repérage aux documents les plus pertinents. De ce fait, il est juste de placer les éléments de précision et l'utilisateur tout au bas de la hiérarchie, c'est-à-dire parmi les éléments de connaissance non utilisés par E4 ou n'ayant que très peu d'influence sur le processus de décision de choix de termes.

Quant à l'exactitude, elle semble refaire surface de manière implicite pour la distinction des DE1 et DE2. En effet, E4 insiste sur la notion de "cœur du document" qui renvoie au thème central représenté par les DE1<sup>1</sup>. Ces termes sont couverts amplement par

---

<sup>1</sup> Rappelons que l'utilisation d'un nouveau thésaurus a altéré la distribution habituelle des descripteurs dans les deux champs principaux. Ainsi, bien que E4 ait identifié le cœur du document pour les documents

le document et proviennent essentiellement du titre et des grands-titres de la table des matières. De ce fait, l'exactitude se situe au bas de la hiérarchie au même titre que les éléments très peu mentionnés ou non utilisés puisqu'elle n'exerce aucune influence pour le choix de descripteurs.

## 2.6 Évaluation comparée des indexats

Ajoutons à cette caractérisation du *satisficing* quelques données issues de l'évaluation comparée des indexats. Puisque ces données sont présentées pour la première fois, nous examinerons et commenterons dans un premier temps, les listes de descripteurs modifiées et les taux de cohérence associés. Ensuite, nous nous attarderons à la valeur convergente des commentaires des indexeurs en fonction de la caractérisation établie du *satisficing* de chaque indexeur.

Rappelons que les entrevues pour l'évaluation comparée des indexats se sont déroulées environ deux mois après l'indexation et consistaient à recueillir les commentaires des indexeurs sur les listes produites par les autres indexeurs ainsi que sur leurs propres listes pour les documents indexés avec le thésaurus Envirodoq, soit les documents *Étang aéré* et *Biogaz*. À l'issue de cette évaluation, les indexeurs pouvaient modifier leurs listes de descripteurs pour ces mêmes documents.

### 2.6.1 Listes de descripteurs modifiées et taux de cohérence interindexeurs

En comparant les nouvelles listes (Tableau XVII, Tableau XIX) aux indexats initiaux (Annexe A), il semble que les indexeurs modifient leurs listes de manière à ajuster le nombre de DE en fonction de la moyenne de DE attribués par l'ensemble des indexeurs. Précisons que les descripteurs qui ont été ajoutés sont représentés par les descripteurs soulignés, les erreurs sont indiquées en italique et la dernière section au bas du tableau présente les descripteurs qui ont été retirés.

---

*Compostage et Énergie*, elle n'a pas ordonné ses descripteurs selon cette distinction, mais plutôt selon l'ordre de consultation du thésaurus.

Ainsi, les représentations succinctes de E1 et E3 pour le document *Étang aéré* deviennent plus élaborées par l'ajout de 2 DE pour E1 et 3 DE pour E3 (Tableau XVII), tandis que E4 réduit à 11 DE sa représentation initiale de 21 DE. E2 n'apporte aucune modification à sa liste initiale qui contient 12 DE.

Tableau XVII. Modification des listes de descripteurs produites pour le document *Étude de sites d'implantation d'étangs aérés*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (12) Descripteurs (12)</b>	<b>Tous les champs (12) Descripteurs (12)</b>	<b>Tous les champs (8) Descripteurs (8)</b>	<b>Tous les champs (11) Descripteurs (11)</b>
<b>Principaux (3)</b> étang aéré <u>emplacement</u> traitement eau usée	<b>Principaux (5)</b> emplacement étang aéré traitement eau usée municipalité évaluation	<b>Principaux (3)</b> évaluation emplacement étang aéré	<b>Principaux (4)</b> étang aéré emplacement évaluation traitement eau usée
<b>Secondaires (1)</b> <u>station épuration</u>	<b>Secondaires (3)</b> répercussions environnementales station épuration coût	<b>Secondaires (2)</b> traitement eau usée <u>station épuration</u>	<b>Secondaires (3)</b> <i>usine épuration</i> <u>répercussions environnementales</u> coût
<b>Descripteurs Géo. (8)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord Lufty Méandre Boysfarm Monfette	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord	<b>Descripteurs Géo. (3)</b> Piedmont <u>St-Sauveur-des-Monts</u> <u>St-Sauveur-Paroisse</u>	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord
<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>
<b>Descripteurs retirés</b>	<b>Descripteurs retirés</b>	<b>Descripteurs retirés</b>	<b>Descripteurs retirés</b> puits contamination habitation topographie superficie zonage dommage vent expropriation zone inondable design

*Note.* Les descripteurs ajoutés sont soulignés. Les termes en italique indiquent une erreur.

Dans sa liste originale, E3 ne retient qu'un seul DE<sub>géo</sub> tandis que les autres indexeurs en retiennent tous un minimum de 4 DE<sub>géo</sub>. Lors des modifications, E3 ajoute 2 DE<sub>géo</sub>. Après avoir consulté les autres indexats, E4 réduit son nombre de DE<sub>2</sub> de 10 à 3 et retire tous ses candidats descripteurs, ce qui rend sa liste comparable aux autres puisque aucun autre indexeur n'avait inscrit de candidats descripteurs pour ce document.

Tableau XVIII. Modification des listes de descripteurs produites pour le document *Élimination du biogaz. Sélection d'une technologie et évaluation des impacts*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
Tous les champs (12)	Tous les champs (9)	Tous les champs (8)	Tous les champs (12)
Descripteurs (8)	Descripteurs (7)	Descripteurs (7)	Descripteurs (11)
Principaux (3) élimination déchet enfouissement sanitaire <u>pollution</u>	Principaux (4) <u>enfouissement sanitaire</u> <u>élimination</u> <u>gaz</u> répercussions environnementales	Principaux (4) technologie élimination répercussions environnementales enfouissement sanitaire	Principaux (5) dépotoir élimination gaz technologie <u>répercussions environnementales</u>
Secondaires (2) <u>dépotoir</u> informatique	Secondaires (2) <u>qualité air</u> échantillonnage	Secondaires (2) qualité air bruit	Secondaires (5) enfouissement sanitaire incinération qualité air échantillonnage bruit
Descripteurs Géo. (3) carrière Miron CUM Montréal	Descripteurs Géo. (1) <u>Montréal</u>	Descripteurs Géo. (1) Montréal	Descripteurs Géo. (1) Montréal
Candidats descripteurs (4) biogaz <u>méthane</u> torchère ISCST	Candidats descripteurs (2) <u>biogaz</u> caractérisation	Candidats descripteurs (1) biogaz	Candidats descripteurs (1) biogaz
Descripteurs retirés pollution bruit logiciel	Descripteurs retirés élimination déchet pollution air technologie gaz carbonique ville Montréal <i>gestion déchet</i>	Descripteurs retirés	Descripteurs retirés étude impact analyse chimique effet serre odeur circulation avion législation règlement cogénération méthane torchère flamme visible

Note. Les descripteurs ajoutés sont soulignés. Les termes en italique indiquent une erreur.

Pour le document *Biogaz* (Tableau XVIII), E4 réduit considérablement son nombre de DE en le faisant passer de 22 à 12. E2 et E3 conservent le même nombre de DE, tandis que la liste de E1 passe de 11 DE à 12. Il est cependant important de préciser que E2 a retiré les DE *élimination déchet*, *pollution air*, *technologie*, *gaz carbonique* et le terme *gestion déchet*, les remplaçant par *enfouissement sanitaire*, *élimination*, *gaz*, *qualité air* et le candidat DE *biogaz*. Il s'agit d'une modification considérable de la représentation initiale qui doit obéir à des considérations de nature plus subtile qu'un simple ajustement quantitatif. D'autre part, soulignons que E3 n'a apporté aucune modification à sa liste.

Ces résultats complémentaires démontrent un plus grand accord lorsque les indexeurs consultent les indexats produits par leurs pairs. En effet, comme l'indique le Tableau XIX, la cohérence interindexeurs augmente de 26 % pour le document *Étang aéré* et de 23 % pour le document *Biogaz*.

Tableau XIX. Taux de cohérence interindexeurs pour la modification des listes de descripteurs des documents *Étang aéré* et *Biogaz*.

Documents	Taux de cohérence interindexeurs	
	Indexations originales	Indexations modifiées
<i>Étang aéré</i>	47 %	73 %
<i>Biogaz</i>	28 %	51 %

Est-ce possible que des indexeurs experts arrivent à un consensus pour représenter un document? Sur quelle base repose ce plus grand accord? Peut-on supposer qu'ils arrivent à juger une indexation meilleure que les autres? Quels seraient alors les éléments de connaissance utilisés pour effectuer une telle évaluation? Et, utilisent-ils les mêmes éléments pour reconnaître une indexation comme la plus satisfaisante?

### 2.6.2 Commentaires des indexeurs sur l'ensemble des indexats

Ajoutons maintenant à la caractérisation du *satisficing* quelques énoncés provenant des entrevues recueillies lors de l'évaluation comparée des listes de descripteurs qui viennent confirmer la priorité de plusieurs éléments pour chaque indexeur. Le fait que les indexeurs reprennent relativement le même type d'argumentation en entrevue confère une certaine validité à la priorité des éléments identifiée à partir des protocoles concomitants et consécutifs.

## E1

On remarque que les éléments usager et capacité de discrimination, prioritaires dans le processus de E1, sont évoqués régulièrement lors de l'entrevue. Par exemple, l'énoncé suivant démontre la cohérence de son argumentation puisqu'elle évoque l'usager en entrevue pour justifier le choix de *logiciel*, lequel avait été évoqué en expérimentation.

**E1 Entrevue/ Commentaires sur sa propre indexation du document Biogaz.**

J'ai mis *informatique*, mais j'ai mis *logiciel* quand même parce qu'*informatique* c'est très large. Je me suis dit, les chercheurs pourraient chercher ça (*ce document*) avec *logiciel*.

Sa prise en compte constante de l'usager apparaît également parmi ses commentaires des autres listes. Dans les énoncés suivants, la prise en compte de l'usager est évoquée pour l'ajout potentiel d'un terme ou pour expliquer le choix de termes d'un autre indexeur.

**E1 Entrevue/ Commentaires sur sa propre indexation du document Étang aéré.**

Je suis sûre qu'on pourrait ajouter des choses à cela (*à sa liste de termes*), des synonymes auxquels les gens penseraient qui ne sont pas nécessairement les mots du titre.

Rappelons que E1 n'a retenu que deux descripteurs (DE1, DE2) pour ce document.

**E1 Entrevue/ Commentaire sur l'indexation du document Étang aéré par E2.**

Peut-être voulait-elle faire un lien logique entre les termes principaux comme *étang aéré* et *traitement eau usée*? En ajoutant *emplacement* ça spécifie. On se met toujours à la place du chercheur, quelqu'un aurait pu chercher *étang aéré*, *traitement eau usée* avec *emplacement* ou avec *municipalité* ou avec *évaluation*. C'est sûr que personne ne va chercher *emplacement* ou *municipalité* tout seul.

On remarque également que la capacité de discrimination des termes est utilisée à nouveau en entrevue pour souligner son approbation ou son désaccord pour le choix de certains termes des listes des autres indexeurs.

**E1 Entrevue/ Commentaire sur l'indexation du document Étang aéré par E2.**

*Évaluation, répercussions environnementales*, ça va revenir dans tous les documents, c'est pour cela que je ne les mettais pas. *Coût* oui, on n'en parle pas nécessairement dans tous les documents des *coûts*.

**E1 Entrevue/ Commentaire sur l'indexation du document Étang aéré par E4.**

L'indexeur n'a pas choisi cela (*tous les termes de sa liste*) pour rien mais dans une étude environnementale, une fois qu'on en a vu quelques-unes, on sait que ça va parler de *zonage*, *superficie* et des différents effets que ça peut avoir.

E2

Les éléments de connaissance exactitude et usager ressortent de manière plus marquée dans l'entrevue de E2. On remarque en effet que E2 reprend les mêmes arguments d'adéquation du sens et d'ampleur du traitement pour juger ses propres descripteurs et ceux des autres indexeurs, comme le soulignent les énoncés suivants:

**E2 Entrevue/ Commentaires sur sa propre indexation du document Biogaz.**

Je n'avais pas *biogaz*, mais je ne suis pas certaine du *gaz carbonique* encore. J'ai peut-être manqué d'information sur ce que c'est vraiment *biogaz*.

*Répercussions environnementales*, c'est bon, je l'aurais mis là aussi. Parce que ce n'est pas une *étude d'impact*, c'est plus les *répercussions environnementales*.

**E2 Entrevue/ Commentaire sur l'indexation du document Étang aéré par E4.**

*Avion, circulation, effet de serre, bruit*. C'est trop pour une description de quelque chose qui est si peu traité dans le document. *Règlement, législation*, je ne sais pas. Probablement qu'elle est rentrée dans les détails de la table des matières encore. Il n'en parle même pas une page parce qu'il y a deux sujets dans la même page. Je ne l'aurais pas mis, ce n'est pas révélateur.

En plus de son souci constant pour l'exactitude, l'usager est également évoqué dans son jugement des listes des autres indexeurs. L'exemple souligne l'exactitude en rapport à l'adéquation du sens entre *habitation* et la notion d'emplacement ou de lieu physique, ainsi que la satisfaction de l'usager en fonction de l'ampleur du traitement des termes *contamination, zonage, superficie*. Implicitement, E2 ne semble pas d'accord avec la majorité des DE2 de la liste de E4 puisqu'ils ne constituent qu'une mince partie du document.

**E2 Entrevue/ Commentaire sur l'indexation du document Étang aéré par E4.**

Trop de DE2. L'usager va se perdre. Pour un si petit document en plus.

*Habitation*, peut-être pour vouloir faire ressortir l'emplacement, le lieux physique. Je ne sais pas si *habitation* est employé dans ces cas-là.

*Contamination*, est-ce qu'ils en parlent? Je ne pense pas. Lorsqu'on met un sujet habituellement, il faut avoir du volume sur ce sujet, un certain nombre de pages. *Zonage, superficie*, non. L'usager va être déçu en voyant comment on traite de *superficie, topographie*, etc.

Par ailleurs, on remarque qu'un indexeur peut transposer sa propre appréhension d'un élément pour expliquer le choix de termes d'un autre indexeur. Ainsi, E2 suppose que E3 a retenu le terme *bruit* en fonction de l'ampleur du traitement alors que E3 a retenu ce terme en fonction de la rareté de cette notion pour le fonds documentaire du MENVIQ.

**E2 Entrevue/ Commentaire sur l'indexation du document *Étang aéré* par E3.**

*Bruit.* Je trouve ça curieux, mais il doit en parler en quelque part. (E2 regarde la table des matières) Il y a des pages, j'aurai peut-être dû le mettre. Probablement pour ça aussi qu'elle a mis *bruit*, parce qu'il était traité plus (*que les autres aspects environnementaux*).

**E3**

Le propos central est l'élément de connaissance privilégié qui ressort des commentaires de E3 en entrevue. En effet, les énoncés suivants témoignent de sa préoccupation constante du propos central lorsqu'elle évalue les listes des autres indexeurs.

**E3 Entrevue/ Commentaire sur l'indexation du document *Biogaz* par E1.**

Je ne sais pas pourquoi elle a mis *élimination déchet*. Elle s'est fiée probablement à l'en-tête *Centre de Tri et d'Élimination des Déchets*. Parce que toute la cause de ce document, c'est *l'élimination des déchets*, mais ce n'est pas de ça dont on parle.

**E3 Entrevue/ Commentaire sur l'indexation du document *Étang aéré* par E2.**

*Station épuration* c'est lié, ce n'est pas mauvais parce que *traitement des eaux usées par étangs aérés* cela implique *station épuration* même si le sujet du document ce n'est pas *station épuration*. *Rivière du Nord*, je ne sais pas. Peut-être que si je regardais encore le document plus attentivement je l'ajouterais aussi, quoique le document ne parle pas de la rivière.

**E3 Entrevue/ Commentaire sur l'indexation du document *Étang aéré* par E4.**

Il a voulu absolument ne rien perdre du document. Essayer de sortir tout ce qu'il avait dans le document, sans avoir en tête l'importance ou est-ce ça le sujet du document. Mais ce n'est pas le sujet du document. Tout cela entre dans évaluation des sites. L'indexeur a dû regarder les titres de chapitre même les plus petits.

**E4**

L'entrevue de E4 renferme moins de commentaires révélateurs des éléments privilégiés par cet indexeur comparativement aux entrevues des autres indexeurs. Toutefois, plusieurs remarques de E4 se rapportent au nombre de termes retenus par les différents indexeurs, ce qui renvoie de façon indirecte à une prise en compte de l'exhaustivité.

**E4 Entrevue / Commentaires sur l'ensemble des indexations du document *Étang aéré*.**

Sur E1: Rapide, laisse des choses de côté.

Sur E2: Balancée, un équilibre dans les termes qui sont là. Il y a des descripteurs, mais pas trop.

Sur E3: Pas assez poussée, bien que les descripteurs utilisés sont tous très importants.

**E4 Entrevue / Commentaires sur l'ensemble des indexations du document *Biogaz*.**

Sur E1 et E3: Un peu superficielles.

Sur E2: C'est bon, mais il manque des descripteurs.

Mais peut-on associer ce type de remarques au souci constant et prioritaire pour l'exhaustivité identifié dans les protocoles de E4? Rappelons que pour E4 l'application de l'exhaustivité consiste principalement en un relevé exhaustif de tous les termes présents dans les parties les plus informatives du document, sans évaluation de la pertinence des termes. Il est possible que ses remarques reposent sur la comparaison du nombre de termes de ses listes avec le nombre de termes des listes des autres indexeurs. En effet, parmi les énoncés précédents, E4 juge que l'indexation de E2 (la plus exhaustive après la sienne) est équilibrée et balancée, tandis que celle de E3 ne contient pas suffisamment de descripteurs. Malgré le fait qu'elle reconnaît que les descripteurs de E3 sont très importants, sa critique principale se rapporte à l'exhaustivité.

Soulignons qu'*a posteriori*, E4 elle-même trouve ses listes trop exhaustives et réduit considérablement le nombre de descripteurs lors de la réindexation, constituant une liste presque identique à celle de E2 pour le document *Étang aéré*, et une liste qui semble combiner les termes de E2 et E3 pour le document *Biogaz*. En d'autres termes, E4 semble ajuster ses nombres de descripteurs en fonction du nombre de descripteurs des listes des autres indexeurs, ce qui remet l'exhaustivité au premier plan de ses considérations.

Bien que les entrevues permettent de confirmer la prise en compte prioritaire de plusieurs éléments de connaissance par les indexeurs, les verbalisations associées à l'évaluation comparée des indexats des documents *Étang aéré* et *Biogaz* ne fournissent presque aucun indice sur les motivations des indexeurs pour le changement de descripteurs de leurs listes. Parfois il est fait mention d'un élément de connaissance pour appuyer le choix d'un nouveau descripteur, mais dans la plupart des cas, les commentaires des indexeurs se limitent au strict choix ou rejet des descripteurs. Ainsi, nous ne nous attarderons pas à cet aspect des données puisque toute interprétation relèverait d'une déduction intuitive. On doit noter cependant que les indexeurs modifient leur liste en choisissant de nouveaux descripteurs à partir de l'ensemble des descripteurs des quatre listes et ne consultent pas le thésaurus. Cette démarche témoigne-t-elle d'une intention de réajuster leur liste en visant un plus grand consensus interindexeurs? Le peu d'indices contenus dans les verbalisations associées à l'évaluation comparée des produits et le cadre informel de cette tâche ne permettent pas de vérifier cette supposition.

## 2.7 Portraits cognitifs des quatre indexeurs.

Le Tableau XX regroupe l'ensemble des éléments de connaissance privilégiés et non privilégiés pour chaque indexeur. Cette représentation hiérarchique fournit un portrait global de chaque indexeur et une comparaison interindexeurs. La première section présente les éléments privilégiés en ordre d'importance et celle du bas comprend les éléments qui semblent n'avoir peu ou pas d'influence lors des choix de termes et descripteurs. L'ordre de présentation des éléments non privilégiés ne correspond pas à une hiérarchie.

En somme, cette hiérarchie permet de répondre de manière plus précise aux questions suivantes : Qu'est-ce qu'un terme *indexable*? et, comment se définit une indexation *satisfaisante* pour chaque indexeur?

Tableau XX. Hiérarchie d'importance des éléments de connaissance pour chaque indexeur.

E1	E2	E3	E4
<b>Éléments privilégiés</b>	<b>Éléments privilégiés</b>	<b>Éléments privilégiés</b>	<b>Éléments privilégiés</b>
Usager Capacité de discrimination Exactitude Actualité Rareté	Exactitude Exhaustivité (facettes) Usager	Propos central Exactitude Précision Rareté	Exhaustivité Actualité Rareté Nouveauté
<b>Éléments non privilégiés</b>	<b>Éléments non privilégiés</b>	<b>Éléments non privilégiés</b>	<b>Éléments non privilégiés</b>
Exhaustivité Précision	Précision Rareté Actualité	Exhaustivité Usager Actualité	Exactitude Précision Usager

En d'autres termes, cette caractérisation démontre que la résolution de l'indexation répond à un *satisficing* multiéléments. La pertinence d'une information (notion, terme, descripteur) pour la représentation du document est évaluée en fonction d'un ensemble d'éléments de connaissance variant d'un indexeur à l'autre. De plus, un point crucial permettant de distinguer la stratégie d'indexation particulière à chaque indexeur réside dans l'importance relative accordée à chacun de ces éléments lors de la prise de décision.

Il est donc intéressant de réexaminer et de comparer les listes d'indexation à la lumière de ces différentes combinaisons hiérarchiques d'éléments et de mettre en évidence ce que ces profils d'indexation permettent d'expliquer dans les variations de choix de termes.

Réexaminons tout d'abord les listes de descripteurs produites pour les documents *Étang aéré* et *Biogaz* (Annexe A). Notons que la caractérisation de chaque indexeur est valable pour l'indexation des quatre documents et ne se limite pas exclusivement à l'indexation de ces deux documents.

La prédominance de l'exhaustivité dans la procédure d'indexation de E4 associée à la non prise en compte de l'exactitude, de la précision et de l'usager explique le nombre élevé de descripteurs en comparaison aux nombres de descripteurs retenus par les autres indexeurs, et ce, pour l'ensemble de ses listes. En effet, E4 tient à n'omettre aucune notion qui passerait sous silence au repérage malgré le peu de pertinence de plusieurs d'entre elles. En conséquence, E4 est le seul indexeur à retenir les termes suivants (descripteurs ou candidats descripteurs) pour le document *Étang aéré*: *puits, contamination, habitation, topographie, superficie, zonage, dommage, vent, expropriation, zone inondable, design*. Pour la même raison, ses autres listes comportent également un nombre élevé de descripteurs que l'on ne retrouve pas dans les listes des autres indexeurs. De plus, sa prise en compte des aspects d'actualité et des notions plus rarement ou nouvellement abordées en environnement contribue à l'exhaustivité particulière de ses listes.

En somme, l'application de l'exhaustivité sans autre condition constitue le trait distinctif majeur de E4 et rend compte de la plupart des variations de descripteurs avec les autres indexeurs.

À l'opposé de E4, la procédure de E1 est dominée par le souci de répondre aux besoins des usagers en plus d'éviter la représentation de notions n'ayant qu'une faible capacité de discrimination. Ceci explique qu'elle ait rejeté plusieurs termes retenus par d'autres indexeurs tels que, *évaluation, répercussions environnementales* pour le document *Étang aéré*. De plus, E1 ne mise pas sur l'exhaustivité mais plutôt sur l'exactitude, autre élément qui oppose E1 à E4. En ce sens, E1 limite sa représentation à quelques DE1 jugés

exacts plutôt que de garnir sa représentation de termes additionnels dont elle n'est pas certaine de l'exactitude et qui pourraient créer de fausses pistes au repérage. Par ailleurs, sa prise en compte prioritaire et particulière de l'utilisateur explique son nombre élevé de DE géo. En effet, E1 retient tous les noms des sites (*Lufty, Boysfarm, etc.* pour le document *Étang aéré*) afin de fournir un accès à l'utilisateur qui n'est pas familiarisé avec le vocabulaire du domaine de l'environnement.

En fait, la combinaison des éléments prioritaires usager et capacité de discrimination rend compte d'une part, du nombre relativement restreint de descripteurs principaux et secondaires des listes de E1 en comparaison aux listes des autres indexeurs. D'autre part, cette combinaison rend compte également du nombre élevé de descripteurs géographiques et de candidats descripteurs de l'ensemble de ses listes.

Ainsi, pour l'indexation du document *Biogaz*, l'utilisateur est l'élément à la base du choix du descripteur *informatique*, même s'il n'est pas précis ou exact pour représenter la notion du document. Les candidats reliés *ISCST* et *logiciel*, ainsi que les descripteurs géographiques *CUM* et *carrière Miron* ont été retenus également pour répondre à l'utilisateur. Notons que ces termes ne se retrouvent que sur la liste de E1. Il ressort alors que la prise en compte unique de l'utilisateur par E1 se reflète directement par un choix de termes différents.

En somme, cette considération unique de l'utilisateur et de la capacité de discrimination constitue le trait distinctif de E1.

Une première explication de la variation de choix de termes de E1 et E4 provient en grande partie de leur prise en compte opposée des éléments usager et exhaustivité. On remarque en effet que l'utilisateur est le premier élément de la hiérarchie de E1 alors qu'il occupe le dernier rang chez E4, tandis que l'exhaustivité apparaît en tête de la hiérarchie de E4 et se retrouve parmi les éléments non privilégiés par E1.

L'exhaustivité et l'utilisateur sont également des éléments privilégiés de la procédure de E2, toutefois, l'application qu'elle en fait permet d'expliquer plusieurs différences de choix de termes avec E1 et E4.

Premièrement, l'exhaustivité de E2 se rapporte à la représentation des différentes facettes (ou aspects) suffisamment couvertes par le document. Il s'agit en fait de la combinaison prioritaire exhaustivité, exactitude, alors que E4 ne mise en priorité que sur l'exhaustivité. C'est ce qui explique que E2 a retenu moins de descripteurs que E4 et plus de descripteurs que E1 et E3 qui ne privilégient pas l'exhaustivité. D'ailleurs, plusieurs termes rares ou d'actualité retenus par E4 ont été rejetés par E2 puisqu'ils ne représentaient pas une facette suffisamment traitée dans le document. Notons que l'importance différente accordée aux éléments actualité et rareté témoignent également de certaines variations des listes de E2 et E4.

La prise en compte de l'utilisateur par E2 est différente de celle de E1 en plusieurs points. Premièrement, E2 ne retient pas de termes pour un type d'utilisateurs spécifique tel que le spécialiste ou le néophyte comme le fait E1. Elle tente plutôt de satisfaire l'ensemble des utilisateurs par une indexation exhaustive et exacte en évaluant globalement sa liste de termes en fonction de la satisfaction des utilisateurs. Parfois, elle sélectionne certains termes plus fréquemment utilisés par les utilisateurs (par exemple, *gestion déchets* et *échantillonnage* pour le document *Biogaz*), sans toutefois négliger les conditions d'exactitude et d'exhaustivité. Par ailleurs, la différence de la position hiérarchique occupée par l'utilisateur et l'exactitude indique également une opposition entre E1 et E2 justifiant plusieurs différences de choix de termes. En effet, E1 accorde la priorité à l'utilisateur pour le choix de ses termes alors que E2 privilégie une représentation exacte.

Cette différence d'appréhension et de priorité attribuée à l'élément utilisateur rend compte de plusieurs variations de choix de termes. Ainsi, E1 juge certains termes primordiaux pour répondre aux demandes potentielles des spécialistes de l'environnement ainsi qu'aux néophytes, tandis que E2 considère ces mêmes termes inadéquats: *Lufty*, *Méandre*, *Boysfarm*, *Monfette* pour le document *Étang aéré*; *informatique*, *carrière Miron*, *CUM*, *ISCST*, *logiciel* pour le document *Biogaz*; *agricoles*, *déchets* pour le document *Compostage*. Notons par ailleurs, que E1 est le seul indexeur à retenir ces termes. Pour sa part, E2 ne retient aucun terme exclusivement pour l'utilisateur puisque les termes doivent répondre en priorité aux éléments d'exactitude et d'exhaustivité, c'est-à-dire correspondre à une facette du document. Cependant, l'utilisateur est évoqué conjointement aux éléments prioritaires pour le choix de plusieurs descripteurs.

Finalement, il ressort que la représentation exacte des diverses facettes constitue le trait le plus distinctif de la stratégie de E2 en comparaison aux autres indexeurs. C'est donc principalement la combinaison exactitude et exhaustivité qui caractérise la nature des choix de E2.

La procédure de E3 se distingue des autres indexeurs en trois points majeurs. Premièrement, elle mise en priorité sur une représentation du propos central du document, ne sélectionnant que très rarement des aspects plus secondaires. Plusieurs des termes secondaires évacués de sa représentation finale sont retenus par les autres indexeurs qui privilégient plutôt une représentation exhaustive, exacte ou répondant aux divers besoins des usagers. En effet, les listes de E3 ne se distinguent pas par un choix de termes particuliers puisque la majorité de ses descripteurs ont également été retenus par l'un ou l'autre des indexeurs. En fait, seul le descripteur *biodégradation* et le candidat *compostage* sont uniques aux listes de E3. La concision de ses listes est donc tributaire de la priorité accordée au propos central. Lors de l'entrevue, tous les indexeurs critiquent d'ailleurs la concision de ses listes mais concèdent que les termes retenus sont très importants.

Deuxièmement, E3 est le seul indexeur à privilégier la précision, ce qui entraîne parfois le rejet de termes qui ne peuvent être traduits adéquatement, ou précisément, par le vocabulaire du thésaurus, et plus particulièrement avec le thésaurus Infoterra. Par exemple, les notions de *valorisation agricole*, de *technologie* et d'*économie* du document *Compostage* ne sont pas traduites en descripteurs, faute de descripteurs suffisamment précis et exacts. Par contre, E3 doit faire des compromis de précision pour représenter plusieurs notions essentielles. Ainsi, le terme *compostage* est traduit par le descripteur *compost et engrais naturels*, *boues résiduaires* est traduit par *déchets liquides et système d'épuration des eaux usées*, et l'aspect urbain par *zones urbaines*. On remarque que E3 est le seul indexeur à inscrire en candidat descripteur le terme précis *compostage* qui représente la notion essentielle de ce document.

Finalement, le dernier trait distinctif de E3 provient de l'ordonnement de ses descripteurs de manière à former un énoncé significatif, c'est-à-dire une phrase résumant le propos du document. Entre autres, ceci explique qu'elle inscrit *évaluation* comme premier

descripteur de sa liste, le document pouvant être résumé par l'énoncé "*évaluation des emplacements pour les étangs aérés à Piedmont*". Notons qu'en entrevue, les indexeurs critiquent la pertinence de mettre ce terme vague en premier puisque cela n'indique pas le propos premier du document.

En résumé, la combinaison propos central, exactitude et précision caractérise le processus d'indexation de E3, ce qui rend compte de la nature des termes choisis et explique le nombre restreint de descripteurs de ses listes.

Cependant, nous devons revenir sur cette concision pour démontrer qu'elle est bien attribuable aux éléments précédemment évoqués et non due à une indexation superficielle ou hâtive, basée principalement sur une représentation des termes du titre. De plus, cette démonstration contribue également à souligner la pertinence de l'utilisation d'un deuxième thésaurus non connu des indexeurs. En effet, l'utilisation du thésaurus Infoterra a permis de mettre en évidence les éléments prioritaires dans le processus de E3, ce que l'utilisation d'Envirodoq ne faisait pas ressortir aussi clairement puisque, connaissant très bien le vocabulaire de ce thésaurus, E3 pouvait choisir directement ses descripteurs sans recourir au langage documentaire. En conséquence, cet automatisme ne laissait que quelques traces d'une prise en compte réelle de la précision et de l'exactitude pour le choix des descripteurs. Dans ce cas particulier, une façon de pouvoir confirmer que les éléments mentionnés ont réellement été appliqués consiste à vérifier leur utilisation lorsque le processus est plus apparent. Ainsi, rappelons que pour l'indexation des documents effectuée à l'aide du thésaurus Infoterra, E3 consulte de manière exhaustive les différentes sections du thésaurus (sommaire des catégories et sous catégories, liste alphabétique et liste par catégorie). Soulignons, par ailleurs, que E3 est le seul indexeur à faire une telle exploration d'Infoterra. Cette consultation en profondeur du thésaurus semble refléter sa préoccupation première de produire une indexation précise et exacte, et ce, quel que soit le langage documentaire utilisé.

## DISCUSSION GÉNÉRALE

Dans cette partie, nous discuterons des résultats en termes plus généraux afin de faire ressortir ce qu'ils permettent d'expliquer face à la problématique du manque de cohérence interindexeurs et ce qu'ils mettent en lumière sur la tâche d'indexation. Nous discuterons également de certaines limites de la présente recherche ainsi que des améliorations à apporter afin que des travaux futurs puissent conduire à une analyse plus complète du travail des indexeurs.

### *Satisficing et hiérarchisation des éléments de connaissance*

Nous avons avancé l'hypothèse que les normes de l'indexation ont été assimilées de manière personnelle par les indexeurs experts, et qu'en conséquence, ils ont développé leur propre méthode pour produire une indexation satisfaisante. Selon cette hypothèse, l'explication des divergences interindexeurs proviendrait d'une prise en compte différente des éléments de connaissance impliqués dans la réalisation de la tâche. Ainsi, en cherchant à identifier les éléments de connaissance utilisés et à expliquer les variations interindexeurs, l'analyse des résultats a permis de dégager différents profils d'indexeurs basés sur une utilisation hiérarchisée d'ensembles d'éléments de connaissance.

En fait, les prises de décisions des indexeurs semblent reposer sur un ensemble restreint d'éléments de connaissance qui, appliqués à la résolution de l'indexation, constituent le *satisficing* autant pour l'évaluation d'un terme que pour l'évaluation du produit global. Ce *satisficing* est un composite de deux, trois ou quatre éléments de connaissance privilégiés et pondérés différemment selon l'indexeur. Il semble que l'atteinte de cette satisfaction représente l'aboutissement d'une négociation constante entre les multiples contraintes de la tâche et l'application des éléments de connaissance privilégiés.

En effet, nos résultats permettent de rendre compte que l'un des problèmes de la prise en compte de ces éléments de connaissance provient de l'impossibilité de maximiser tous les éléments simultanément, d'où la formation d'une stratégie personnelle. Par exemple, un indexeur misant sur l'exhaustivité pourra laisser de côté un peu d'exactitude ou de précision. À l'inverse, une application rigoureuse de la précision aura une conséquence directe sur

l'exhaustivité. En ce sens, il existerait une sorte de tension interéléments intrinsèque à la tâche à laquelle les indexeurs répondent en élaborant leur propre stratégie de résolution.

Ainsi, pour E1 une indexation satisfaisante doit contenir des termes qui répondent aux besoins d'une vaste clientèle, qui ont une forte capacité de discrimination et inversement, qui ne représentent pas des termes n'ayant qu'une faible capacité de discrimination. De plus, ces termes doivent représenter les notions avec exactitude, et E1 mise également sur les notions d'actualité et celles plus inusités. Pour E2, une indexation satisfaisante doit comporter des termes qui représentent le contenu du document de manière exacte et exhaustive et qui répondent aux besoins de la clientèle. Pour E3, une indexation satisfaisante est constituée de termes qui représentent le propos central du document avec exactitude et précision, ainsi que des notions plus rarement traitées dans les documents du centre de documentation. Quant à E4, sa définition d'une indexation satisfaisante est constituée d'un nombre de termes permettant de représenter le contenu du document de manière très exhaustive, sans oublier les notions se rapportant à l'actualité et à la rareté.

Par ailleurs, cette hiérarchisation dans l'importance accordée à un ensemble particulier d'éléments de connaissance semble également fournir quelques indices en rapport à ce que constitue l'*indexabilité* d'un terme (Bates, 1998; Beghtol, 1986; Milstead, 1994; Todd, 1992). En effet, il est possible d'établir un parallèle entre critère de satisfaction et *indexabilité* d'un terme qui pourrait être associée aux éléments de connaissance privilégiés par chaque indexeur. Donc, selon cette caractérisation, l'*indexabilité* d'un terme ou d'une notion se rattacherait à l'évaluation qu'en fait l'indexeur en relation aux éléments privilégiés plutôt qu'uniquement à la présentation particulière de l'information (*propos*) dans le document.

*Types d'approches: orientée usager, orientée document et orientée langage documentaire*

Nos résultats permettent également de situer les profils des indexeurs selon les deux approches de l'indexation, soit l'indexation orientée-document et orientée-usager (Albrechtsen, 1993; Fidel, 1994). Rappelons qu'en théorie comme en pratique une

indexation devrait souscrire aux deux approches. Cependant, l'ordre hiérarchique attribué aux éléments de connaissance fait ressortir l'une ou l'autre approche du processus global d'indexation de chaque indexeur.

Ce qui singularise nos résultats n'est pas le fait que l'on puisse "classer" les indexeurs en fonction de cette polarité, mais plutôt que l'on fournisse un meilleur portrait de ce que signifie concrètement ces approches, c'est-à-dire, la façon dont s'articule, dans le processus même d'indexation, la combinaison particulière d'éléments de connaissance privilégiés par un indexeur permettant de caractériser ces approches. Ainsi, bien que deux indexeurs puissent avoir un profil général similaire, le niveau de nos analyses permet de mettre en lumière quelques distinctions plus fines associées à chaque type de profil. Aussi, l'on voit qu'une préoccupation pour l'exactitude n'est pas gage uniquement d'une approche orientée-document, mais peut témoigner d'un souci d'une représentation fidèle du contenu pour favoriser un meilleur repérage souscrivant ainsi aux deux approches. L'identification de la hiérarchisation dans l'utilisation des divers éléments de connaissance permet d'associer une combinaison particulière d'éléments à l'une ou l'autre de ces approches.

De plus, selon nos analyses, il semble que l'approche orientée-usager couvre probablement plus large que l'utilisateur lui-même et englobe potentiellement les aspects plus fonctionnels, tels que le fonds documentaire, la suggestion de candidats descripteurs pour l'amélioration et la mise à jour du vocabulaire du thésaurus et la prise en compte des dimensions actualité, nouveauté, rareté, etc.

Ainsi soulignons brièvement l'approche privilégiée par chaque indexeur. Avec l'utilisateur comme élément pivot, E1 possède une approche plus orientée-usager. Au pôle opposé, on retrouve E3 qui, privilégiant la représentation du propos central du document, adopte une approche orientée-document. E2 occupe une position mitoyenne puisqu'elle privilégie une représentation exacte de l'ensemble des facettes du document, tout en évaluant leur pertinence en fonction de la satisfaction de l'utilisateur et en s'attardant aux termes d'indexation les plus fréquemment demandés par les utilisateurs. Il semble donc s'agir d'une double approche orientée à la fois document et usager, avec une tendance légèrement plus marquée pour le document selon la hiérarchie des éléments privilégiés.

Par ailleurs, le processus d'indexation de E4 introduit une nouvelle approche, bien que de prime abord cet indexeur semble privilégier une approche orientée-document. En effet, l'élément d'exhaustivité prioritaire et la longueur de ses listes témoignent d'une représentation de la plupart des notions présentées dans le document. Cependant, la sélection systématique de tous les descripteurs qui permettent de représenter un terme de la liste initiale sans évaluation de la pertinence (exactitude, précision, utilité pour l'utilisateur, etc.) dénote une représentation, du moins une sélection de descripteurs, qui est guidée principalement par les possibilités offertes par le thésaurus. Ainsi, bien que E4 représente les termes centraux comme les autres indexeurs, son processus global d'indexation peut être qualifié d'approche orientée-document pour la première étape de l'indexation et orientée-thésaurus lorsqu'il s'agit de la transposition des termes en descripteurs. Ajoutons, que cette procédure s'apparente au principe de surindexation, biais reconnu dans la pratique de l'indexation consistant à noter chaque notion rencontrée sans procéder à une discrimination méticuleuse de l'information réellement *indexable*.

#### *Représentation du propos*

La mise à jour de la hiérarchisation dans l'application des éléments de connaissance renvoie également à la nature du *propos* représenté par les indexeurs. Pour reprendre la terminologie proposée par Beghtol (1986), un document posséderait un *propos* intrinsèque ou extensionnel (*aboutness*) relativement permanent et indépendant de toute utilisation potentielle, et un *propos* intentionnel se rapportant aux raisons pour lesquelles un usager pourrait en faire la demande. Ainsi, le *propos* représenté serait de nature différente selon que le document est analysé en tant qu'unité isolée et finie ou analysé en rapport avec les dimensions extratextuelles (besoins des usagers, fonds documentaire, etc.) liées à la tâche.

Ainsi, les indexeurs de notre expérimentation ont démontré une prise en compte possible de ces concepts théoriques du *propos*, lesquels peuvent être associés dans la pratique aux deux grandes approches orientée-document et orientée-usager, où le *propos* extensionnel s'apparente à l'approche orientée-document alors que le *propos* intentionnel correspond plus à l'approche orientée-usager.

Ceci répond également à l'hypothèse avancée par Milstead (1994) qui supposent que le constat d'incohérence interindexeurs réside dans les perceptions ou interprétations différentes du document. En effet, la tendance à avoir une approche plus orientée-document ou orientée-usager laisse présumer que l'indexeur accorde une importance plus marquée pour la représentation du *propos* extentionnel ou intentionnel.

Les arguments évoqués par les indexeurs pour leurs choix de termes indiquent une distinction relativement claire entre ces deux propos. Le *propos* extentionnel pourrait être associé aux termes centraux, pour lesquels les indexeurs sont généralement en accord, et qui serait associé aux termes issus de la page-titre et des grands-titres de la table des matières. La représentation du *propos* intentionnel pourrait alors être associé à la prise en compte de tous les éléments de connaissance relatifs aux dimensions extratextuelles (usager, fonds documentaire, actualité, etc.).

Ainsi, la caractérisation des profils des indexeurs n'est plus basée uniquement sur les approches orientée-document ou orientée-usager, mais tient compte également des dimensions extratextuelles utilisées par les indexeurs qui couvrent les éléments de connaissance actualité, rareté, capacité de discrimination, nouveauté, etc. La représentation du *propos* du document semble parfois éclatée puisque la préoccupation première de l'indexeur n'est pas de représenter uniquement le contenu du document, mais bien de souligner les aspects qui sont susceptibles d'intéresser les usagers et de représenter ces aspects de manière à ce que différents types d'usagers puissent les repérer aisément. Il arrive également que le *propos* est représenté de façon minimale, c'est-à-dire sans indication de la particularité d'un document par rapport aux autres documents similaires du fonds documentaire. Par exemple, avec la priorité accordée aux éléments usager et capacité de discrimination, les listes de E1 illustrent une représentation axée sur le *propos* intentionnel alors que celles de E3 reflètent plus la représentation du *propos* extentionnel, ayant en priorité l'élément propos central.

Évidemment, il ne s'agit pas ici de réduire la caractérisation des indexeurs à ces quelques notions, mais bien de situer nos résultats en rapport à des préoccupations théoriques soulevées par certains auteurs.

### *Représentation de l'utilisateur*

La dimension usager ressort de nos données comme particulièrement influente sur les choix de termes et descripteurs. Il a été souligné à plusieurs reprises qu'il est primordial de tenir compte des besoins de la clientèle dans le processus d'indexation, malgré que ces besoins soient difficilement prévisibles et varient en fonction des intérêts particuliers de chaque usager.

En examinant la place accordée à l'utilisateur dans le processus réel d'indexation, nous avons dégagé trois types de prises en compte de l'utilisateur, lesquels entraînent un choix de termes différents et conséquemment permettent de rendre compte de certaines variations interindexeurs reliées à cet élément.

Ainsi, nous avons démontré qu'un indexeur qui tient à indexer pour des profils précis d'utilisateurs, en l'occurrence les spécialistes et les néophytes, peut retenir des termes qu'il est seul à considérer pertinents. En fait, la pertinence de ces termes est évaluée uniquement en fonction de l'utilisateur particulier que l'indexeur se représente plutôt qu'en fonction du contenu du document. Un terme peut être sélectionné parce qu'il reflète les préoccupations actuelles du grand public, ce qui en fait un terme d'actualité, sans que l'ampleur du traitement soit un facteur d'évaluation. Il peut également être retenu en descripteur secondaire pour indiquer à l'utilisateur que le document présente cet aspect, mais qu'il n'est pas relié directement au sujet premier du document. Par conséquent, les listes de cet indexeur ressortiront de l'ensemble des listes produites par les autres indexeurs pour ces mêmes documents, plus encore lorsque l'indexeur fait de l'utilisateur l'élément primordial de son processus.

Le second type de prise en compte de l'utilisateur réside dans l'évaluation des termes en fonction de la satisfaction de l'utilisateur. L'indexeur ne semble pas avoir un profil particulier d'utilisateur en tête, mais évalue régulièrement son analyse du point de vue de l'utilisateur. L'influence de cette prise en compte semble se refléter de deux façons sur le choix de termes. Premièrement, l'indexeur a tendance à retenir des termes qu'il croit plus susceptibles d'être utilisés par les utilisateurs. Deuxièmement, l'indexeur choisit des termes qui spécifient l'angle sous lequel un terme déjà retenu est abordé (souvent une notion principale), afin de fournir à l'utilisateur suffisamment d'information pour qu'il puisse décider

si ce document répond à ses attentes. Évidemment, cette dernière prise en compte s'apparente aussi à une indexation plus exhaustive et il est parfois difficile de distinguer clairement l'influence plus marquée de l'un ou l'autre de ces éléments sur le choix final.

En filigrane, on remarque également une attention particulière pour l'utilisateur en post-repérage, c'est-à-dire, une fois que l'utilisateur a repéré la notice. En fait, certains indexeurs semblent ordonner leurs descripteurs afin de fournir des indices sur les liens qui unissent les notions à l'intérieur du document ou sur l'angle de traitement de ces notions, pour que l'utilisateur puisse évaluer et repérer dès ce moment les notices les plus pertinentes en rapport à son besoin. Sans remettre en question la validité réelle d'une telle démarche, soulignons qu'il s'agit d'une stratégie ingénieuse de la part des indexeurs pour tenter de dépasser les contraintes liées au langage documentaire utilisé.

#### *Rôle et influence du langage documentaire*

Ceci nous amène à discuter du rôle de la connaissance du langage documentaire et de son influence sur le processus d'indexation en rapport à une préoccupation théorique émise par Beghtol (1986) à laquelle notre étude répond en partie. Beghtol (1986) a souligné le besoin d'études empiriques afin de démontrer la manière dont les évaluations différentes de la pertinence des termes d'indexation prennent forme lors de l'analyse du propos de documents. Ce qui constitue selon elle, une des causes du manque de cohérence interindexeurs puisque l'indexeur ne pourra faire abstraction de sa connaissance du langage documentaire dès son premier contact avec le document à indexer.

Nos analyses démontrent en effet que les indexeurs ont recours à leur connaissance du langage pour sélectionner plusieurs de leurs termes dès la première étape de l'indexation. Le contenu des feuilles de travail et des listes de descripteurs (Annexe B) illustre ce fait avec évidence puisque plusieurs des termes inscrits sur ces listes correspondent à des descripteurs. L'exemple le plus probant de cette stratégie de sélection directe de descripteurs dès la phase d'extraction provient de la comparaison des listes produites par E3 pour l'indexation des documents *Étang aéré* et *Biogaz*. En effet, rappelons que parmi ses descripteurs finaux, la totalité des descripteurs (cinq termes) ont été extraits directement du document et inscrits, sous leur forme exacte, sur sa feuille de travail pour le document

*Étang aéré* et que des huit descripteurs retenus pour le document *Biogaz*, six étaient présents sur sa feuille de travail.

Sans aucun doute, la connaissance du langage influence l'analyse du contenu du document, les indexeurs devant représenter le document à l'intérieur des possibilités offertes par le langage. Cependant, nos résultats ont révélé que la prise en compte prioritaire de certains éléments de connaissance exerce une influence déterminante sur le choix de descripteurs finaux, ce qui permet de relativiser l'impact du langage sur l'élaboration de la représentation du document.

En effet, bien que les indexeurs aient en tête le vocabulaire du thésaurus lorsqu'ils entreprennent l'analyse d'un document, et qu'en conséquence leur carte mentale ou réseau de relations sémantiques corresponde à ce langage et soit contraint à ce dernier, l'indexeur ne limite pas nécessairement son interprétation en fonction des possibilités du langage.

Ainsi, l'indexeur qui sélectionne des termes en fonction d'un l'élément privilégié, les insérera vraisemblablement dans sa représentation finale et ce, malgré les contraintes imposées par le thésaurus. L'indexeur inscrira le terme préférentiel dans sa liste de descripteurs et proposera aussi un candidat descripteur, fera un compromis d'exactitude ou de précision, multipliera les descripteurs pour représenter une seule notion, ordonnera ses descripteurs de façon à indiquer les relations existantes ou pour préciser un aspect, etc.

À ce titre, rappelons la différence d'utilisation des deux langages par E3 qui démontre, à notre avis, l'importance des éléments privilégiés pour le choix de descripteurs et la stratégie globale d'indexation. Ainsi, E3 qui tient à une représentation précise et exacte, choisit les descripteurs appropriés pour répondre à ces éléments de connaissance. Cette sélection se fait de façon quasi automatique pour les descripteurs extraits d'Envirodoq puisqu'elle connaît ce thésaurus. Pour l'utilisation du thésaurus non familier, Infoterra, E3 doit parcourir les différentes sections du thésaurus pour maintenir la précision et l'exactitude de sa représentation. E3 ne se contente donc pas de sélectionner le premier descripteur jugé valable pour la traduction des termes retenus, mais applique les éléments de précision et d'exactitude en parcourant exhaustivement le vocabulaire qui ne lui est pas familier. Étant le seul indexeur à procéder de cette façon, il apparaît légitime d'attribuer

cette stratégie à sa prise en compte marquée pour les éléments exactitude et précision. D'ailleurs, les verbalisations associées mettent en évidence la préoccupation de E3 pour ces éléments.

Ces différentes constatations illustrent la manière dont les indexeurs se sont appropriés les règles et normes et ont développés des stratégies personnelles leur permettant de contourner plusieurs contraintes des langages et de tenir compte de l'élément ou de l'ensemble d'éléments privilégiés dans leur représentation finale. De plus, l'argumentation sous-jacente aux choix de termes et descripteurs révèle que les indexeurs ne font que très peu de compromis lorsqu'il s'agit de l'élément ou des éléments de connaissance privilégiés. Toutefois, certains compromis demeurent inévitables pour le choix de descripteurs. Il n'est pas rare dans de tels cas, que les verbalisations témoignent de l'insatisfaction des indexeurs à ne pouvoir tenir compte de l'élément privilégié.

Nous devons également ajouter quelques précisions concernant l'utilisation d'un deuxième thésaurus. Les raisons de cette manipulation ont été clairement exposées, au chapitre portant sur la méthodologie, ce choix s'étant révélé pertinent pour mettre à jour les stratégies d'indexation des indexeurs. Cependant, il faut ajouter que les résultats de l'indexation avec Infoterra ne peuvent être considérés comme une réelle indexation puisque, comme nous l'avons constaté, les indexeurs se réfèrent au vocabulaire du thésaurus dès le début de leur analyse. Ainsi, indexer avec un thésaurus ne consiste pas uniquement à consulter un thésaurus, mais implique que l'indexeur a le vocabulaire en tête. On ne peut donc prétendre que les indexeurs auraient retenu les mêmes descripteurs que ceux retenus dans la cas présent, si leur connaissance d'Infoterra avait été équivalente à leur connaissance d'Envirodoq. Rappelons toutefois que les indexeurs ont utilisé les mêmes ensembles d'éléments de connaissance pour évaluer termes et descripteurs quel que soit le thésaurus utilisé.

#### *Complémentarité des verbalisations*

Autre point intéressant. Nous avons souligné à plusieurs reprises le caractère complémentaire des trois types de verbalisations utilisées pour recueillir des données de différents niveaux. Il apparaît clairement, à la lumière des résultats, que ce choix

méthodologique a été judicieux pour cerner avec exactitude la composition du *satisficing* des indexeurs. Peu de recherches ont utilisé ces méthodes, et à notre connaissance, aucune n'avait encore combiné les trois.

### *Contributions et pistes futures*

Certaines des données obtenues laissent entrevoir la possibilité d'enrichir et de raffiner les portraits cognitifs d'indexation identifiés.

En effet, nous avons identifié et hiérarchisé les éléments de connaissance utilisés fréquemment par les indexeurs pour l'analyse des documents. Nous avons traité ces éléments de manière isolée dans un premier temps et en relation les uns avec les autres dans un deuxième temps. De cette dernière analyse, émerge la caractérisation du *satisficing* ou une pondération différente de l'importance accordée par l'indexeur à chaque élément de connaissance évoqué pour le choix de termes et descripteurs ainsi que pour l'évaluation du produit final.

Cependant, certaines décisions semblent liées à des idiosyncrasies ou à des facteurs dont nous n'avons pas tenu compte dans nos analyses puisque les minces traces relevées ne permettaient pas d'en faire une analyse précise. Par exemple, le fait qu'un indexeur juge parfois des termes pertinents en fonction de sa connaissance du monde ou qu'un indexeur s'attarde à quelques notions en fonction de ses intérêts personnels témoigne de cas relativement isolés, mais bel et bien présents. Ainsi, il est possible que certaines différences de choix de descripteurs proviennent de variations de la connaissance générale desquelles nos analyses ne peuvent rendre compte.

Il s'agit évidemment de connaissances qui seront appelées en avant-plan dans la mémoire de l'indexeur lorsqu'il analyse le contenu d'un document. En parcourant le document, l'indexeur ne peut faire abstraction de tout son bagage cognitif et conséquemment des liens s'établissent aussi avec une foule d'éléments n'ayant pas directement rapport avec la tâche précise d'indexation. Cependant, en fonction des objectifs de notre étude, nous avons délibérément centré nos analyses sur les éléments directement liés à la tâche et avons privilégié le choix d'indexeurs possédant un profil des

plus similaire en rapport aux variables intrinsèques à la tâche, soit la connaissance de l'indexation, la connaissance du langage documentaire et la connaissance du domaine.

Ainsi, nos analyses proposent une nouvelle explication à la problématique de la cohérence interindexeurs et les différents ensembles d'éléments utilisés par les indexeurs sont suffisamment documentés dans notre recherche pour être tenus pour valables.

La tâche d'indexation est complexe et nous ne prétendons en aucune façon expliquer tout le phénomène, c'est-à-dire rendre compte de toutes les variables contextuelles et individuelles susceptibles d'entrer en jeu dans l'indexation. Toutefois, ce que nos résultats ont permis de dégager constitue une première caractérisation de l'utilisation réelle des éléments de connaissance reconnus par les normes comme essentiels à la qualité de l'indexation. De plus, nos résultats montrent comment l'application différentielle des paramètres définis par les normes engendre des indexats différents.

Maintenant, si nous voulions poursuivre l'étude du phénomène de la cohérence interindexeurs, quelles améliorations pourraient être apportées à la présente recherche, et quelles pistes mériteraient d'être développées?

Soulignons en premier lieu, qu'un certain flou demeure dans la détermination des différents profils d'indexeur. Nous avons recueilli une trace écrite de deux représentations du contenu d'un document, c'est-à-dire, la liste brouillon de termes susceptibles d'être retenus pour la traduction en descripteurs et la liste finale de descripteurs. Cependant, tout au long du processus, plusieurs termes traités ne sont pas écrits, et bien que nous en conservions une trace dans les protocoles, il est parfois difficile de suivre l'évolution du traitement et les éléments de connaissance associés à ce traitement jusqu'à la décision finale (c'est-à-dire choix ou rejet). La nature même de la tâche rend ce suivi relativement difficile et parfois impraticable. Par exemple, il arrive que durant l'analyse du document, l'indexeur débute le traitement d'un terme, bifurque vers un autre terme sans revenir sur le terme précédent, le traitement de ce dernier pouvant être laissé en suspens. Il devient alors plus difficile pour l'expérimentateur de reconnaître le moment le plus approprié pour questionner l'indexeur sur ce terme, ne sachant pas si l'indexeur traitera à nouveau ce terme

précis. En effet, rappelons que contrairement au problème bien défini tel que la Tour de Hanoï (Figure 2, p.68), l'indexation possède les caractéristiques d'un problème mal défini.

Rappelons que le recueil des verbalisations consécutives s'effectue tout suite après l'indexation des quatre documents, ce qui ne laisse aucune marge à l'expérimentateur pour réviser les bandes vidéo des protocoles concomitants avant de passer à la seconde étape de l'expérimentation<sup>1</sup>. Il est donc possible que les raisons du choix ou rejet de certains termes et descripteurs ne ressortent pas dans les données.

Ainsi, il se peut que certains éléments jugés non appliqués par un indexeur ne soient tout simplement pas visibles dans le processus, s'ils n'ont été évoqués que mentalement par l'indexeur et non explicitement rappelés en verbalisation consécutive.

En somme, nous avons caractérisé le *satisficing* d'indexeurs experts à partir de l'indexation de quatre documents. Mais que trouverait-on si l'on répétait cette expérience avec un plus grand nombre de documents et un plus grand nombre d'indexeurs? Et que pourrait-on améliorer dans la méthode pour permettre un recueil plus exhaustif des éléments de connaissance utilisés dans la tâche d'indexation?

Premièrement, l'indexation d'un plus grand nombre de documents permettrait probablement de confirmer et de stabiliser l'identification du profil de chaque indexeur, ainsi que de raffiner l'analyse de ce profil par la découverte de dimensions supplémentaires ou par une meilleure caractérisation des situations dans lesquelles chaque élément de connaissance est utilisé.

Si l'on faisait indexer plus de documents aux indexeurs ayant participé à notre expérimentation, cela permettrait peut-être de faire ressortir plus clairement l'impact de certains éléments de connaissance moins fréquemment évoqués ou de certaines règles

---

<sup>1</sup> Ces choix méthodologiques ont été exposés en détails au chapitre II, mais rappelons uniquement qu'il était nécessaire de procéder ainsi afin que la trace mnémonique de l'indexation soit présente de la manière la plus complète dans la mémoire des indexeurs.

personnelles, mentionnées en verbalisations, auxquelles l'indexeur ne semble cependant pas avoir recours concrètement pour l'analyse des différents documents.

Par exemple, E2 rapporte s'être fixé une limite arbitraire de six descripteurs par champ (DE1 et DE2). Bien que l'on ait identifié cette règle personnelle, à première vue très déterminante, nous n'avons pu vérifier son application réelle ou son incidence sur ses indexations. En effet, E2 n'a pas eu à retrancher de descripteurs de ces listes n'ayant pas retenu plus de notions que la limite prédéterminée.

Cette règle est surprenante *a priori* puisqu'elle vient fixer une limite quantitative indépendante de l'analyse sémantique. Dans une certaine mesure, cette règle peut se justifier en fonction de la nature des documents généralement indexés au MENVIQ (rapports de recherche, études de faisabilité, etc.), lesquels nécessitent rarement, voir exceptionnellement, plus d'une douzaine de descripteurs (DE1, DE2). Ainsi, l'on pourrait vérifier l'application de cette règle personnelle si, parmi le plus grand nombre de documents à indexer, certains nécessitaient un nombre élevé de descripteurs pour la représentation de leur contenu.

Par ailleurs, l'indexation d'un plus grand nombre de documents permettrait possiblement de raffiner la caractérisation de l'utilisation d'éléments mentionnés moins fréquemment dans les protocoles d'un indexeur, mais qui semblent néanmoins avoir une forte incidence sur le choix de termes ou descripteurs. Ainsi, l'élément de rareté d'une notion privilégié par E3 pourrait faire l'objet d'une attention particulière. Cet élément s'est révélé important pour l'identification du *satisficing* de cet indexeur, puisque E3 semble faire un compromis d'exactitude et de précision au profit de la rareté d'une notion.

À travers l'indexation d'une plus grande quantité de documents, nous pourrions également porter une attention particulière aux éléments de connaissance qui se sont révélés cruciaux dans la caractérisation du *satisficing* des indexeurs et que nous n'avons pas recensés de prime abord. Ainsi, les éléments d'actualité, de rareté, de nouveauté et de capacité de discrimination feraient partie des éléments essentiels dont la prise en compte serait vérifiée. Plus précisément, cette vérification systématique pourrait consister en un

questionnement lors des verbalisations consécutives, semblable à ce que nous avons fait pour l'élément usager.

Rappelons que pour le recueil des verbalisations consécutives, deux types de questionnements ont été utilisés: un premier questionnement se rapportant aux situations ponctuelles issues des protocoles concomitants, et un deuxième questionnement, de niveau plus général, permettant de préciser la manière dont les divers éléments de connaissance sont utilisés dans la stratégie d'indexation de l'indexeur. Ainsi, lorsque l'indexeur mentionne un élément essentiel, tel que l'utilisateur, nous le questionnions alors de manière plus directe sur son utilisation de cet élément.

Dans une étude ultérieure, les indexeurs pourraient être questionnés plus en profondeur sur les éléments de connaissance recensés comme les plus influents sur les choix de termes et descripteurs des indexeurs dans la présente recherche.

Une autre amélioration serait souhaitable pour une éventuelle continuation de cette recherche. Il s'agirait d'apporter quelques modifications à l'étape de l'évaluation comparée des indexats. Rappelons brièvement qu'il s'agissait pour les indexeurs de commenter les listes de descripteurs produites par les autres indexeurs ainsi que leurs propres listes dans le but de provoquer la réflexion de l'indexeur sur les motivations des autres indexeurs pour leurs choix de descripteurs. Ensuite, les indexeurs pouvaient modifier leur liste initiale de descripteurs. L'analyse de ces données a révélé que les indexeurs semblent inférer les stratégies de leurs pairs en fonction des éléments de connaissance qu'eux-mêmes privilégient.

Cependant, le peu d'indices contenus dans ces verbalisations relèguent ces constatations au niveau de suppositions qui pourraient faire l'objet d'une étude plus rigoureuse pour être confirmées. De plus, on ne peut tirer aucune conclusion des commentaires des indexeurs associés aux modifications de leurs indexats.

Une façon d'enrichir le recueil de données pour cette étape pourrait consister à resserrer le questionnement. En effet, l'entrevue se déroulait de manière relativement informelle, ce qui a pu contribuer à la maigreur des données recueillies. Aussi, nous

croyons qu'une entrevue plus dirigée permettrait de faire ressortir plus d'éléments. Par exemple, lors de la modification des listes de descripteurs, l'indexeur pourrait être questionné systématiquement sur chacune de ses modifications. La durée de l'entrevue devrait alors être augmentée.

Par ailleurs, l'entrevue s'effectuait deux mois après le recueil des deux premiers types de protocoles. Serait-il plus judicieux de faire suivre plus rapidement ces deux étapes? Nous voulions que les indexeurs puissent se distancier de leurs propres indexations pour être en mesure de commenter de manière impartiale l'ensemble des listes présentées. Ce laps de temps était peut-être trop important pour que les indexeurs puissent commenter les listes de manière élaborée.

En somme, ces données supplémentaires n'ont pas apporté la contribution espérée pour la confirmation de la caractérisation du *satisficing* des indexeurs et les raisons proviennent, à notre avis, de la méthode utilisée pour le recueil de données.

À quoi pourrait-on s'attendre d'une expérimentation impliquant l'indexation de plusieurs documents par un grand nombre d'indexeurs? Peut-être verrait-on apparaître l'utilisation d'éléments de connaissance supplémentaires, non identifiés à travers l'indexation des quatre documents par nos quatre indexeurs.

L'analyse des stratégies individuelles d'indexation d'un grand nombre d'indexeurs permettrait probablement l'identification d'un nombre limité de profils d'indexeurs. En d'autres termes, nous supposons qu'il ne peut exister une variété infinie de *satisficing* puisque les indexeurs ont recours à un ensemble relativement restreint d'éléments de connaissance déterminants pour l'évaluation des termes et descripteurs. Malgré le fait que d'autres éléments viendraient s'ajouter au nombre d'éléments déjà identifiés, on parviendrait éventuellement à une stabilité dans l'identification du *satisficing*. En somme, une étude plus vaste sur la cohérence interindexeurs permettrait de déterminer les grands portraits cognitifs pour l'ensemble de la population des indexeurs œuvrant dans les centres de documentation spécialisés se servant de thésaurus.

Nos analyses ont permis de démontrer la réelle préoccupation de certains indexeurs pour l'élément usager et l'influence notable de cet élément dans le choix de termes et descripteurs. Ainsi, cette dimension pourrait faire l'objet d'une étude plus systématique qui viserait à identifier la nature des variations interindexeurs selon des profils d'utilisateurs précis. Nous ne tenterons pas ici d'élaborer un plan de recherche pour une telle étude, mais l'ampleur de l'influence de l'utilisateur sur le processus d'indexation des indexeurs justifie à notre avis que davantage d'études se penchent sur la représentation des utilisateurs chez les indexeurs.

Finalement, soulignons qu'il pourrait être intéressant de présenter nos résultats à des groupes d'indexeurs et de recueillir leurs commentaires. Sont-ils conscients de l'ensemble des éléments de connaissance qui s'affrontent constamment et des compromis nécessaires dans le processus d'indexation? Quelles sont les dimensions de la tâche qui les préoccupent consciemment? Aurait-ils des éléments nouveaux à apporter pour la compréhension de la problématique de la cohérence interindexeurs? Ceci enrichirait notre compréhension de la cohérence interindexeurs du point de vue des "praticiens" eux-mêmes.

## CONCLUSION

Les implications de notre recherche se situent à plusieurs niveaux. D'une part, elles jettent une nouvelle lumière sur la problématique de la cohérence interindexeurs, et d'autre part, elles fournissent quelques explications sur l'articulation de plusieurs principes théoriques dans le processus réel d'indexation.

Le cœur de notre contribution se rapporte à l'identification et à la caractérisation du *satisficing* des indexeurs pour expliquer les variations interindexeurs pour le choix de descripteurs.

Premièrement, nos résultats ont révélé l'influence prépondérante de divers éléments de connaissance pour la détermination de l'*indexabilité* d'un terme et la production d'une indexation *satisfaisante*. Ainsi, nous avons identifié que les choix des indexeurs reposent sur l'utilisation d'un ensemble hiérarchisé d'éléments de connaissance variant d'un indexeur à l'autre. Ceci permet d'expliquer la nature des choix des indexeurs et par conséquent de rendre compte des variations interindexeurs. Ainsi, la stratégie propre à chaque indexeur se retrouve en partie dans cette utilisation personnelle d'un ensemble particulier d'éléments de connaissance. Jusqu'à présent, aucune recherche n'avait permis de mettre à jour la manière dont ces éléments de connaissance influencent concrètement les décisions de l'indexeur.

Nos résultats ont également démontré la manière dont se traduit concrètement une approche orientée-document et orientée-usager dans le processus, en plus d'avoir identifié une approche orientée-thésaurus. L'approche privilégiée doit également être nuancée selon la hiérarchisation des éléments de connaissance utilisés par l'indexeur.

Il semblerait que l'indexeur cherchera à contourner les contraintes imposées par le langage documentaire afin d'appliquer l'élément ou l'ensemble d'éléments de connaissance qu'il privilégie, ce qui permet de relativiser le rôle majeur attribué au langage documentaire dans l'élaboration de la représentation du contenu d'un document.

Autre point important, l'usager ressort comme une dimension cruciale pour l'explication des variations interindexeurs. En effet, nous avons identifié trois types de

considérations qui permettent de rendre compte de la nature différente des termes choisis selon que l'indexeur vise à répondre à des profils précis d'utilisateurs (spécialistes et néophytes) ou qu'il vise à répondre globalement à la clientèle. Les listes de l'indexeur qui accorde la priorité à l'élément utilisateur se singularisent par rapport aux listes produites par les autres indexeurs pour un même document.

De plus, nous sommes conscients que pour pouvoir généraliser nos conclusions, il serait nécessaire d'effectuer une expérimentation plus vaste incluant un plus grand nombre d'indexeurs. De même, il serait intéressant de considérer le rôle d'autres facteurs dans la résolution de l'indexation selon notre approche et notre méthode, ce qui permettrait de fournir des portraits cognitifs plus riches des indexeurs.

Nos résultats ne permettent pas de prédire le comportement d'autres indexeurs œuvrant dans un contexte similaire, mais l'ensemble des éléments identifiés pourrait constituer une grille de base pour l'étude du travail d'un tel groupe.

Par ailleurs, nous n'avons pas anticipé l'effet à long terme qu'une telle expérimentation aurait sur le travail des indexeurs. Deux des indexeurs ayant participé à notre recherche ont rapporté être devenus conscients de leur procédure, ce qui les amène à questionner leurs choix de termes et descripteurs lors de l'analyse. Il s'agit d'une contribution imprévue de notre étude qui laisse entrevoir la possibilité d'une application concrète pour la formation des indexeurs ou pour des interventions dans les centres de documentation.

Il s'agirait alors de sensibiliser les indexeurs, qu'ils soient novices ou experts, à l'impact de la prise en compte des divers éléments de connaissance dans la résolution de l'indexation. Ainsi, il serait intéressant de leur exposer les types d'actions intellectuelles et les questionnements auxquels ils sont confrontés ainsi que les pistes potentielles menant à la décision finale, c'est-à-dire, les divers critères sur lesquels ils peuvent baser leur évaluation pour le choix de termes et descripteurs, tels que nous les avons identifiés. Cette conscientisation n'entraînerait pas nécessairement une amélioration de la cohérence interindexeurs, mais pourrait faire ressortir l'importance d'établir des rencontres régulières pour les indexeurs d'un même centre afin qu'ils puissent mettre en commun leurs stratégies

individuelles et qu'en conséquence, ils puissent ajuster leurs processus d'indexation. Ainsi, nous supposons que de tels échanges permettraient probablement une meilleure cohérence, non pour les choix précis de termes, mais pour l'application des divers éléments de connaissance.

Nous laissons le mot de la fin à deux chercheurs (Jacob et Shaw, 1998) qui positionnent notre recherche (David, Giroux, Bertrand-Gastaldy, Lanteigne et Bertrand, 1995)<sup>1</sup> en fonction des études en sciences de l'information. Cet article soulève le fait que notre recherche est importante, non seulement pour son utilisation de méthodes non traditionnelles pour l'examen du travail des indexeurs, mais également pour la formation des futurs indexeurs. À propos de la contribution de notre recherche, ces auteurs ajoutent:

"by investigating not only the cognitive process(es) involved in representational activities but also the impact of the indexer's mental model of a good representation on the final product, it may be possible to provide future indexers with effective domain-based pragmatic frameworks that will enhance consistency of representation in individual knowledge communities." (Jacob et Shaw, 1998, p.170)

---

<sup>1</sup> Il s'agit d'une publication rapportant les résultats d'analyses préliminaires de notre expérimentation.

## RÉFÉRENCES

- Agosto, D. E. (2002). Bounded Rationality and Satisficing in Young People's Web-Based Decision Making. *Journal of the American Society for Information Science and Technology*, 53(1), 16-27.
- Albrechtsen, H. (1993). Subject Analysis and Indexing: from Automated Indexing to Domain Analysis. *The Indexer*, 18(4), 219-224.
- Anzai, Y., Simon, H. A. (1979) The Theory of Learning by Doing. *Psychological Review*, 86(2), 124-140.
- Association française de normalisation (1978). *Principes généraux pour l'indexation des documents (NF Z 47-102 1978)*. (1ère ed.). Paris: AFNOR.
- Bainbridge, L. (1979). Verbal Report as Evidence of the Process Operator's Knowledge. *International Journal of Man-Machine Studies*, 11, 411-436.
- Bardini, T. (1996). Changements et réseaux socio-techniques. De l'inscription à l'affordance. *Réseaux*, 76, 126-155.
- Bates, M. J. (1998). Indexing and Access for Digital Libraries and the Internet : Human, Database, and Domain Factors. *Journal of the American Society for Information Science*, 49(13), 1185-1205.
- Bates, M. J. (2002). The Cascade of Interactions in the Digital Library Interface. *Information Processing and Management*, 38, 381-400.
- Beghtol, C. (1986). Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents. *Journal of Documentation*, 42(2), 84-113.
- Belkin, N. J. (1984). Cognitive Models and Information Transfer. *Social Science Information Studies*, 4, 111-129.
- Bell, C., Jones, K. P. (1979). Computerized Indexing Need not Be Impossible. *The Indexer*, 11(3), 132-135.
- Bertrand, A. (1993). *Compréhension et Catégorisation dans une activité complexe: L'indexation de document scientifiques*. Thèse de Doctorat Nouveau Régime, non publiée, Université de Toulouse-Le Mirail, Toulouse, France.
- Bertrand, A., Cellier, J.-M. (1995). Psychological Approach to Indexing: Effects of the Operator's Expertise upon Indexing Behaviour. *Journal of Information Science*, 21(6), 459-472.

- Bertrand, A., Cellier, J.-M., Giroux, L. (1994). Expertise et stratégies dans une activité d'indexation de documents scientifiques. *Le Travail Humain*, 57(1), 25-51.
- Bertrand, A., Cellier, J.-M., Giroux, L. (1996). Expertise and Strategies for the Identification of the Main Ideas in Document Indexing. *Applied Cognitive Psychology*.
- Bertrand-Gastaldy, S. (1986). De quelques éléments à considérer avant de choisir un niveau d'analyse ou un langage documentaire. *Documentation et Bibliothèques, janvier-juin*, 3-23.
- Bertrand-Gastaldy, S. (1990). *Les différents types de traitement analytique et de langages documentaires*. Montréal: Faculté des arts et des sciences, École de bibliothéconomie et des sciences de l'information, Université de Montréal.
- Bertrand-Gastaldy, S., Giroux, L., Lanteigne, D., David, C. (1994). Les produits et processus cognitifs de l'indexation humaine. *ICO Québec, printemps 1994*, 6(1-2), 29-40.
- Branch, J. L. (2000). The Trouble with Think Alouds : Generating Data Using Concurrent Verbal Protocols. Présentation à l'ACSI 2000 : *Les dimensions d'une science de l'information globale*, Université d'Alberta, Edmonton, Canada, 28-30 mai 2000. En ligne : <http://www.slis.ualberta.ca/cais2000/branch.htm>.
- Breuleux, A. (1990). *L'élaboration et l'exécution de plans dans une tâche de rédaction*. Thèse de doctorat, non publiée. Université de Montréal, Montréal.
- British Standard Institution. (1984). *Examining documents, Determining their Subjects and Selecting Indexing Terms (BS 6529 ; 1984)*. (1st ed.). Londres: British Standard Institution.
- Bureau canadien des archivistes (1992). *L'indexation par sujet en archivistique. Rapport du Groupe de travail sur l'indexation par sujet*. Ottawa: Bureau canadien des archivistes.
- Burgin, R. (1991). The Effect of Indexing Exhaustivity on Retrieval Performance. *Information Processing and Management*, 27(6), 623-628.
- Caverni, J. P. (1988). La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif. In J. P. Caverni, C. Bastien, P. Mendelsohn, et G. Tiberghien (Éds.). *Psychologie cognitive. Modèles et méthodes* (pp.253-273). Grenoble : Presse Universitaire de Grenoble.

- Centre de documentation et de renseignements du MENVIQ. (1993). *Thésaurus Envirodoq*. Québec: Ministère de l'Environnement du Québec.
- Chaumier, J. (1980). *Travail et méthodes du/de la documentaliste*. Paris: Entreprise moderne d'édition.
- Chaumier, J. (1988). *Le traitement linguistique de l'information*. Paris: Entreprise moderne d'édition.
- Chen, C., Czerwinski, M., Macredie, R. (2000). Individual Differences in Virtual Environments- Introduction and Overview. *Journal of the American Society for Information Science*, 51(6), 499-507.
- Courrier, Y. (1976). Analyse et langage documentaires. *Documentaliste*, 13(5-6, septembre-décembre), 178-189.
- Crystal, D. (1995). Is There Anybody There? *The Indexer*, 19(3), 153-154.
- David, C., Giroux, L., Bertrand-Gastaldy, S., Lanteigne, D., Bertrand, A. (1995). Indexing as Problem Solving : A Cognitive Approach to Consistency. *Proceedings of the 58th Annual Meeting of the American Society for Information Science*, 32, 49-55.
- Deffner, G., Rhenius, D. (1985). Experimental Tests of the Ericsson/Simon Model of Concurrent Thinking Aloud. In G. d'Ydewalle (Éd.) *Cognition, Information Processing, and Motivation*, (pp. 495-504). Amsterdam: Elsevier Science.
- Desbois, B. (1985). *Étude de la possibilité d'utiliser les énergies renouvelables, (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal*. Mémoire de maîtrise, non publié. Université du Québec à Montréal, Montréal.
- Dufresne, A. (2002). Les filtres à la communication sur Internet. In J. Lajoie et É. Guichard (Éds.), *Odyssée Internet* (pp. 79-95). Montréal: PUQ - Presses de l'Université du Québec.
- Dupourqué, L. (2002). Le catalogage des sites Web: vers une organisation structurée du Web. *Documentation et bibliothèques*, 48 (2), 34-68.
- Educational Resources Information Center. (1980). *ERIC Processing Manual : Rules and Guidelines for the Acquisition, Selection, and Technical Processing of Documents and Journal Articles by the Various Components of the ERIC network*. Washington, D.C. : Educational Resources Information Center.
- Elio, P., Scharf, P. B. (1990). Modeling Novice-to-Expert Shifts in Problem-Solving Strategy and Knowledge Organization. *Cognitive Science*, 14, 579-639.

- Endres-Niggemeyer, B. (1989). Content Analysis: A Special Case of Text Comprehension. In S. Koskiala and R. Laundo (Eds.). *Information, Knowledge, Evolution. Proceedings of the 44th FID Congress*, Helsinki, 1988, (pp.103-112). Amsterdam : North- Holland.
- Endres-Niggemeyer, B. (1990). A Procedural Model of Abstracting, and Some Ideas for its Implementation. *Proceedings of the second international congress on Terminology and Knowledge Engineering, TKE'90, 1*, 230-243.
- Endres-Niggemeyer, B. (1994). Summarising Text for Intelligent Communication. *Knowledge Organisation*, 21(4), 213-223.
- Ericsson, K. A., Simon, H. A. (1980). Verbal Report as Data. *Psychological Review*, 87(3), 215-251.
- Ericsson, K. A., Simon, H. A. (1984). *Protocol Analysis : Verbal Report as Data*. Cambridge, MA.: MIT Press.
- Eysenck, M. W. (1984). *A Handbook of Cognitive Psychology*. London: Lawrence Erlbaum Associates.
- Farrow, J. (1991). A Cognitive Process Model of Document Indexing. *Journal of Documentation*, 47(2), 149-166.
- Farrow, J. (1994). Indexing as a Cognitive Process. *Encyclopedia of Library and Information Science*, 53 (supp. 16), 155-171.
- Farrow, J. (1995). All in the Mind: Concept Analysis in Indexing. *The Indexer*, 19(4), 243-247.
- Farrow, J. (1996). Propositional Analysis and Macrorules in Indexing. *Library Review*, 45(1), 6-15.
- Fidel, R. (1994). User-Centered Indexing. *Journal of the American Society for Information Science*, 45(8), 572-576.
- Fitzgerald, M. A., Galloway, C. (2001). Relevance Judging, Evaluation, and Decision Making in Virtual Libraries : A Descriptive Study. *Journal of the American Society for Information Science and Technology*, 52(12), 989-1010.
- Fortin, C., Rousseau, R. (1989). *Psychologie cognitive : une approche de traitement de l'information*. Québec: Presses de l'Université du Québec.
- Gardner, H. (1993). *Histoire de la révolution cognitive. La nouvelle science de l'esprit*. Paris: Payot.

- Giroux, L., Larochelle, S. (1987). *L'ergonomie cognitive des systèmes informatiques: état de la question et pistes de recherche*. Centre canadien de recherche sur l'informatisation du travail, Ministère des Communications du Canada.  
Montréal: Université de Montréal et Centre canadien de recherche sur l'informatisation au travail.
- Green, R. (1995). Topical Relevance Relationship. 1. Why Topic Matching Fails. *Journal of the American Society for Information Science*, 46(9), 646-653.
- Greeno, J. G., Simon, H. A. (1988). Problem Solving and Reasoning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce (Eds.). *Steven's Handbook of Experimental Psychology, Vol.21* (pp. 589-672). N.Y.: Wiley.
- Hirsh, S. G. (1999). Children's Relevance Criteria and Information Seeking on Electronic Resources. *Journal of the American Society for Information Science*, 50(14), 1265-1283.
- Hjørland, B. (2001). Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content... and Relevance. *Journal of the American Society for Information Science and Technology*, 52(9), 774-778.
- Hoc, J.-M. (1984). La verbalisation provoquée pour l'étude du fonctionnement cognitif. *Psychologie Française*, 29(3-4), 231-234.
- Hoc, J.-M., Leplat, J. (1983). Evaluation of Different Modalities of Verbalization in a Sorting Task. *International Journal of Man-Machine Studies*, 18, 283-306.
- Holyoak, K. (1990). Problem solving. In D.N. Osherson and E.E. Smith (Eds.). *Thinking: an Invitation to Cognitive Science, Vol. 3* (pp. 117-146). Cambridge, MA.: MIT Press.
- Hutchins, E. (1995). How a Cockpit Remembers its speeds. *Cognitive Science*, 19, 265-288.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA : MIT Press.
- Hutchins, W. J. (1975). *Languages of Indexing and Classification. A Linguistic Study of Structures and Functions*. Stevenage, England : Peter Peregrinus.
- Hutchins, W. J. (1977). On the Problem of Aboutness in Document Analysis. *Journal of Informatics*, 1, 17-35.
- Hutchins, W. J. (1987). Summarization: Some Problems and Methods. Presented at the *Informatics*, 9. In Spark Jones, K. (Ed.) *Meaning: the frontier of informatics, Informatics 9* (pp. 151-173). London: ASLIB.

- Jacob, E. K., Shaw, D. (1998). Sociocognitive Perspectives on Representation. *Annual Review of Information Science and Technology (ARIST)*, 33, 131-185.
- Janes, J. W. (1991). Relevance Judgments and the Incremental Presentation of Document Representations. *Information Processing and Management*, 27(6), 629-646.
- Jones, K. P. (1976). Documentation Notes: Towards a Theory of Indexing. *Journal of Documentation*, 32(2), 118-125.
- Jones, K. P. (1983). How do we Index? : a Report of Some ASLIB Informatics Group Activity. *Journal of Documentation*, 39(1), 1-23.
- Kintsch, W., van Dijk, T. A. (1978). Towards a Model of Text Comprehension and Production. *Psychological Review*, 85(5), 363-394.
- Kolodner, J. L. (1983). Towards an Understanding of the Role of Experience in the Evolution from Novice to Expert. *International Journal of Man-Machine Studies*, 19, 497-518.
- Lancaster, F. W. (1998). *Indexing and Abstracting in Theory and Practice*. Champaign, IL: Graduate School of Library and Information Science, University of Illinois.
- Lancaster, F.W. (2003). Do Indexing and Abstracting have a Future? *Anales de Documentación*, 6, 137-144. En ligne: <http://www.um.es/fccd/anales/ad06/ad0609.pdf>
- Lawrence, S., Giles, C. (1998). Searching the World Wide Web. *Science*, 5360(28), 98-100.
- Leonard, L. E. (1977). *Inter-indexer Consistency Studies, 1954-1975: a Review of the Literature and Summary of the Study Results*. (Occasional papers, no. 131). University of Illinois : Graduate School of Information Science.
- Leplat, J. (1980). *La psychologie ergonomique*. Paris: PUF, coll. Que sais-je?
- Lunin, L. F., Fidel, R. (1994). Introduction and Overview. *Journal of the American Society for Information Science*, 45(8), 570-571.
- MacDougall, S. (1996). Rethinking Indexing: the Impact of the Internet. *The Australian Library Journal*, november, 281-285.
- Maglaughlin, K. L., Sonnenwald, D. H. (2002). User Perspectives on Relevance Criteria : A Comparison among Relevant, Partially Relevant, and Not-Relevant Judgments. *Journal of the American Society for Information Science and Technology*, 53(5), 327-342.
- Marchionini, G. (1995). User-centered Methods for Library Interface Design. *SIGOIS Bulletin. Special Issue on Digital Libraries*, 16(2), 34-36.

- Markey, K. (1984). Inter-indexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials. *Library and Information Sciences Research*, 6, 155-177.
- Millerand, F. (2003) *L'appropriation du courrier électronique en tant que technologie cognitive chez les enseignants chercheurs universitaires. Vers l'émergence d'une culture numérique?* Thèse de doctorat, non publiée. Université de Montréal, Montréal.
- Milstead, J. L. (1994). Needs for Research in Indexing. *Journal of the American Society for Information Science*, 45(8), 577-582.
- Milstead, J. L. (2000). Classification and Representation. *Bulletin of the American Society for Information Science*, 26(2), 13-15.
- Ministère de l'Environnement. (1984). *Compostage des boues résiduelles urbaines*. Tours, France: Ministère de l'environnement.
- Mizzaro, S. (1997). Relevance : The Whole History. *Journal of the American Society for Information Science*, 48(9), 810-832.
- Morris, R. C. T. (1994). Toward a User-Centered Information Service. *Journal of the American Society for Information Science*, 45(1), 20-30.
- Newell, A. (1980). Reasoning, Problem Solving, and Decision Processes: The Problem Space as a Fundamental Category. In R. S. Nickerson (Ed.). *Attention and Performance VIII* (pp. 693-718). Hillsdale, New Jersey: LEA Lawrence Erlbaum Associates Publishers.
- Newell, A., Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Organisation internationale de normalisation (1985). *Documentation : Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation (ISO 5963-1985)*. (1ère ed.). Genève: ISO.
- Patch-Neveu, C. (1989). *Réseau québécois d'information sur la communication (RQIC) : Bibliocom / Badicom*. Montréal: Faculté des arts et des sciences, École de bibliothéconomie et des sciences de l'information, Université de Montréal.
- Programme des Nations Unies pour l'Environnement (1990). *Infoterra : Thésaurus des termes relatifs à l'environnement*. (3e éd.). Nairobi, Kenya: PNUE.
- Propp, V. (1968). *Morphologie du conte*. Paris: Seuil.

- Quinn, B. (1994). Recent Theoretical Approaches in Classification and Indexing. *Knowledge Organisation*, 21(3), 140-147.
- Rabasse, P. (1989). *Mise en évidence des sources de variations et des processus cognitifs inhérents à la tâche d'indexation documentaire*. Mémoire de maîtrise, non publié. Paris: Université de Paris-Sud Orsay.
- Reiman, P., Chi, M. T. C. (1989). Human Expertise. In K. J. Gilhooly (Ed.). *Human and Machine Problem Solving* (pp. 161-191). N.Y.: Plenum Press.
- Robertson, S. E. (1979). Between Aboutness and Meaning. Presented at Informatics 5, Oxford. In McCafferty, M. and Gray, K. (Eds.) *The analysis of meaning : Informatics 5*. (pp.202-205). London: Aslib.
- Rolling, L. (1981). Indexing Consistency, Quality and Efficiency. *Information Processing and Management*, 17, 69-76.
- Saàdani, L., Bertrand-Gastaldy, S. (2000). La représentation des connaissances d'un domaine dans Internet. *Documentation et bibliothèques* 46(1), janvier-mars, 27-42.
- Salton, G., Buckley, C. (1990). Improving Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41(4), 288-297.
- Saracevic, T. (1991). Individual Differences in Organizing, Searching and Retrieving Information, *ASIS '91 Systems Understanding People. Proceedings of the 54th Annual Meeting of the American Society for Information Science*, 28, 82-86. Washington, D.C. 27-31 october.
- Shobowale, G. (1998). SGML, XML and the Document-Centered Approach to Electronic Medical Records. *Bulletin of the American Society for Information Science*, 25(1), 7-10.
- Sievert, M. C., Andrews, M. J. (1991). Indexing Consistency in Information Science Abstracts. *Journal of the American Society for Information Science*, 42(1), 1-6.
- Simon, H. A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, 4, 181-201.
- Simon, H. A. (1985). Information-Processing Theory of Human Problem Solving. In A. M. Aitkenhead and J. M. Slack (Eds.). *Issues in Cognitive Modeling* (pp. 253-278). London: Lawrence Erlbaum Associates.
- Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, 41, 1-19.

- Société Québécoise d'Assainissement des eaux (1990). *Étude de sites d'implantation d'étangs aérés*. Laval, Québec: SQAÉ.
- Soergel, D. (1994) Indexing and Retrieval Performance : The Logical Evidence. *Journal of the American Society for Information Science*, 45(8), 589-599.
- Sparck Jones, K. (1973). Does Indexing Exhaustivity Matter? *Journal of the American Society for Information Science*, 24(5), 313-316.
- Suchman, L. (1987). *Plans and situated action. The problem of human-machine communication*. Cambridge, MA : Cambridge University Press.
- Suchman, L. (1993). Response to Vera and Simon's Situated Action : A Symbolic Interpretation. *Cognitive Science*, 17(1), 71-75.
- Suchman, L. (1995). no title. *SIGOIS Bulletin. Special Issue on Digital Libraries*, 16(2), 15-16.
- Tecresult (1992). *Élimination du biogaz. Sélection d'une technologie et évaluation des impacts*. (Rapport préliminaire). Montréal: Ville de Montréal, Services des travaux publics, Module Génie de l'environnement.
- Todd, R. J. (1992). Academic Indexing: What's it All About? *The Indexer*, 18(2), 101-104.
- Tomonori, G. (1983). Cognitive Structure in Human Indexing Process. *Library and Information Science*, 21, 209-226.
- Tonta, Y. (1991). A Study of Indexing Consistency between Library of Congress and British Library Catalogers. *Library Resources and Technical Services*, 35(2), 177-188.
- van Dijk, T. A., Kintsch, W. (1983). *Strategies of Discourse Comprehension*. London: Academic Press.
- Vera, A. H., Simon, H. A. (1993). Situated Action : A Symbolic Interpretation. *Cognitive Science*, 17(1), 7-48.
- Wang, P., Soergel, D. (1998). Cognitive Model of Document Use during a Research Project. Study 1. Document Selection. *Journal of the American Society for Information Science*, 49(2), 115-133.
- Wellisch, H. H. (1994). Some Comments on Subject Analysis and Indexing. *The Indexer*, 19(2), 131-132.
- Wickelgren, W. (1974). *How to Solve Problems. Elements of a Theory of Problems and Problem Solving*. San Francisco: Freeman & Company.

- Zhang, X. (2002). Collaborative Relevance Judgment : A Group Consensus Method for Evaluating User Search Performance. *Journal of the American Society for Information Science and Technology*, 53(3), 220-231.
- Zunde, P., Dexter, M. E. (1969). Factors Affecting Indexing Performance, *Proceedings of the 32nd Annual Meeting of the American Society for Information Science* (pp. 313-322). Westport, CT: Greenwood.

## Annexe A

Listes de descripteurs produites par les indexeurs  
(feuilles détachables)

Tableau X. Listes de descripteurs produites pour le document *Étude de sites d'implantation d'étangs aérés*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (10)</b>	<b>Tous les champs (12)</b>	<b>Tous les champs (5)</b>	<b>Tous les champs (21)</b>
<b>Descripteurs (10)</b> <b>Principaux (2)</b> étang aéré traitement eau usée	<b>Descripteurs (12)</b> <b>Principaux (5)</b> emplacement étang aéré traitement eau usée municipalité évaluation	<b>Descripteurs (5)</b> <b>Principaux (3)</b> évaluation emplacement étang aéré	<b>Descripteurs (18)</b> <b>Principaux (4)</b> étang aéré emplacement évaluation traitement eau usée
<b>Secondaires</b>	<b>Secondaires (3)</b> répercussions environnementales station épuration coût	<b>Secondaires (1)</b> traitement eau usée	<b>Secondaires (10)</b> <i>usine épuration</i> <sup>1</sup> puits contamination habitation topographie superficie zonage coût dommage vent
<b>Descripteurs Géo. (8)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord Lufty Méandre Boysfarm Monfette	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord	<b>Descripteurs Géo. (1)</b> Piedmont	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord
<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs (3)</b> expropriation zone inondable design

<sup>1</sup> Un terme en italique indique une erreur de la part de l'indexeur. Soit qu'il s'agit d'un terme non préférentiel, c'est-à-dire d'un terme inscrit dans le thésaurus qui n'est pas un descripteur, ou soit que le terme inscrit n'est pas dans le thésaurus. Ce terme n'est donc pas pris en compte dans le calcul de la cohérence interindexeurs.

Tableau XI. Listes de descripteurs produites pour le document *Élimination du biogaz. Sélection d'une technologie et évaluation des impacts.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Tous les champs (11)</b>	<b>Tous les champs (9)</b>	<b>Tous les champs (8)</b>	<b>Tous les champs (22)</b>
<b>Descripteurs (7)</b> <b>Principaux (3)</b> élimination déchet enfouissement sanitaire pollution bruit	<b>Descripteurs (8)</b> <b>Principaux (5)</b> élimination déchet pollution air technologie répercussions environnementales gaz carbonique <b>Secondaires (2)</b> <i>gestion déchet</i> échantillonnage	<b>Descripteurs (7)</b> <b>Principaux (4)</b> technologie élimination répercussions environnementales enfouissement sanitaire  <b>Secondaires (2)</b> qualité air bruit	<b>Descripteurs (18)</b> <b>Principaux (5)</b> dépotoir élimination gaz technologie étude impact <b>Secondaires (12)</b> enfouissement sanitaire incinération qualité air analyse chimique échantillonnage effet serre odeur bruit circulation avion législation règlement <b>Descripteurs Géo. (1)</b> Montréal
<b>Descripteurs Géo. (3)</b> carrière Miron CUM Montréal	<b>Descripteurs Géo. (1)</b> ville Montréal	<b>Descripteurs Géo. (1)</b> Montréal	<b>Descripteurs Géo. (1)</b> Montréal
<b>Candidats descripteurs (4)</b> biogaz ISCST logiciel torchère	<b>Candidats descripteurs (1)</b> caractérisation	<b>Candidats descripteurs (1)</b> biogaz	<b>Candidats descripteurs (4)</b> biogaz cogénération méthane torchère flamme visible

Tableau XII. Listes des descripteurs produites pour le document *Compostage des boues résiduaires urbaines*.

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<p>Tous les champs (6)</p> <p><b>Descripteurs (6)</b>  <b>Principaux (5)</b>            épuration des eaux usées, système d'épuration des eaux usées</p> <p><b>Secondaires</b></p> <p><b>Descripteurs Géo. (1)</b>            France</p> <p><b>Candidats descripteurs</b></p>	<p>Tous les champs (6)</p> <p><b>Descripteurs (6)</b>  <b>Principaux (3)</b>            composts et engrais naturels zones urbaines système d'épuration des eaux usées</p> <p><b>Secondaires (2)</b>            techniques de conversion des déchets économique, développement</p> <p><b>Descripteurs Géo. (1)</b>            France</p> <p><b>Candidats descripteurs</b></p>	<p>Tous les champs (7)</p> <p><b>Descripteurs (6)</b>  <b>Principaux (5)</b>            déchets liquides composts et engrais naturels zone urbaine système d'épuration des eaux usées biodégradation</p> <p><b>Secondaires</b></p> <p><b>Descripteurs Géo. (1)</b>            France</p> <p><b>Candidats descripteurs (1)</b>            compostage</p>	<p>Tous les champs (8)</p> <p><b>Descripteurs (8)</b>  <b>Principaux (6)</b>            usine de traitement des eaux d'égout technique de conversion des déchets récupération des déchets composts et engrais naturels amélioration du sol recyclage</p> <p><b>Secondaires (2)</b>            terres agricoles déchets liquides</p> <p><b>Descripteurs Géo.</b></p> <p><b>Candidats descripteurs</b></p>

Tableau XIII. Listes des descripteurs produites pour le document *Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire et de la biomasse) au village de Ndia Goureye au Sénégal.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
Tous les champs (9)	Tous les champs (12)	Tous les champs (6)	Tous les champs (22)
<b>Descripteurs (8)</b> <b>Principaux (5)</b> énergie solaire énergie de la biomasse énergie éolienne pays en développement ressources renouvelables	<b>Descripteurs (12)</b> <b>Principaux (5)</b> énergie éolienne énergie solaire énergie de la biomasse pays en développement utilisation des ressources locales	<b>Descripteurs (6)</b> <b>Principaux (4)</b> énergie solaire énergie éolienne bouse comme combustible pays en développement	<b>Descripteurs (22)</b> <b>Principaux (12)</b> climatologie désertification zones arides sources d'énergie non-polluante ressources renouvelables énergie solaire énergie éolienne bouse comme combustible puits <i>approvisionnement en eau</i> zones rurales énergie électrique <b>Secondaires (8)</b> précipitations vents changement climatique sécheresse utilisation des ressources locales éclairage biogaz facteurs socio-économiques
<b>Secondaires</b> approvisionnement en eau dans les zones rurales éclairage biogaz <i>fumier</i> sources d'énergies	<b>Secondaires (5)</b> approvisionnement en eau dans les zones rurales éclairage biogaz <i>fumier</i> sources d'énergies	<b>Secondaires (1)</b> source d'énergie non-polluante	changement climatique sécheresse utilisation des ressources locales éclairage biogaz facteurs socio-économiques
<b>Descripteurs Géo. (3)</b> Sénégal Ndia Goureye Sahel <b>Candidats descripteurs (1)</b> rapport de recherche	<b>Descripteurs Géo. (2)</b> Ndia Goureye Sénégal <b>Candidats descripteurs</b>	<b>Descripteurs Géo. (1)</b> Sénégal <b>Candidats descripteurs</b>	<b>Descripteurs Géo. (2)</b> Sénégal Sahel <b>Candidats descripteurs</b>

## Annexe B

Feuilles brouillon et listes de descripteurs produites par les indexeurs  
(feuilles détachables)

Feuilles brouillon et listes de descripteurs pour le document :  
*Étude de sites d'implantation d'étangs aérés.*

<b>EXPERT 1</b>	<b>EXPERT 2</b>	<b>EXPERT 3</b>	<b>EXPERT 4</b>
<b>Feuille brouillon</b> traitement des eaux usées Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord biofiltration Société SQAÉ CEDEGER usine d'épuration usine étangs aérés implantation dézonage	<b>Feuille brouillon</b> étang aéré sites d'implantation traitement eaux usées Rivière du Nord CEDEGER Ité Gendron Lefebvre inc. réhabilitation usine épuration emplacement coût biofiltration	<b>Feuille brouillon</b> étangs aérés Piedmont emplacement traitement eau usée St-Sauveur-des-Monts St-Sauveur-Paroisse station épuration évaluation	<b>Feuille brouillon</b> étang aéré site design évaluation contraintes grille d'évaluation eaux usées Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord usine d'épuration zonage visibilité puits contamination vent habitations contraintes physiques contraintes techniques coût dommages expropriation zone inondable superficie topographie zone résidentielle
<b>Tous les champs (10)</b>	<b>Tous les champs (12)</b>	<b>Tous les champs (5)</b>	<b>Tous les champs (21)</b>
<b>Descripteurs (10)</b> <b>Principaux (2)</b> étang aéré traitement eau usée  <b>Secondaires</b>	<b>Descripteurs (12)</b> <b>Principaux (5)</b> emplacement étang aéré traitement eau usée municipalité évaluation <b>Secondaires (3)</b> répercussions environnementales station épuration coût	<b>Descripteurs (5)</b> <b>Principaux (3)</b> évaluation emplacement étang aéré  <b>Secondaires (1)</b> traitement eau usée	<b>Descripteurs (18)</b> <b>Principaux (4)</b> étang aéré emplacement évaluation traitement eau usée  <b>Secondaires (10)</b> <i>usine épuration</i> puits contamination habitation topographie superficie zonage coût dommage vent
<b>Descripteurs Géo. (8)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord Lufty Méandre Boysfarm Monfette	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord	<b>Descripteurs Géo. (1)</b> Piedmont	<b>Descripteurs Géo. (4)</b> Piedmont St-Sauveur-des-Monts St-Sauveur-Paroisse Rivière du Nord
<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs (3)</b> expropriation zone inondable design

Feuilles brouillon et listes de descripteurs pour le document :  
*Élimination du biogaz. Sélection d'une technologie et évaluation des impacts.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Feuille brouillon</b> destruction du biogaz site d'enfouissement CTED composition destruction Montréal qualité de l'air caractérisation échantillonnage élimination torchères à flamme invisible CUM carrière Miron son bruit COV (composés organiques volatils) génération ISCST modélisation compresseurs revue de littérature logiciel	<b>Feuille brouillon</b> déchet tri gaz échappement technologie impact répercussions environnementales élimination site d'enfouissement enfouissement sanitaire destruction	<b>Feuille brouillon</b> élimination déchet biogaz technologie répercussion environnementales enfouissement sanitaire dépotoir centre de tri élimination qualité air	<b>Feuille brouillon</b> biogaz élimination déchet technologie évaluation impact enfouissement déchet effet de serre pollution atmosphérique incinération retombée atmosphérique émission législation bruit odeurs aéroport circulation aérienne échantillonnage débit composition combustion torchères à flamme visible qualité air anaérobie méthane cogénération captage
<b>Tous les champs (11)</b>	<b>Tous les champs (9)</b>	<b>Tous les champs (8)</b>	<b>Tous les champs (22)</b>
<b>Descripteurs (7)</b> <b>Principaux (3)</b> élimination déchet enfouissement sanitaire pollution bruit  <b>Secondaires (1)</b> informatique	<b>Descripteurs (8)</b> <b>Principaux (5)</b> élimination déchet pollution air technologie répercussions environnementales gaz carbonique <b>Secondaires (2)</b> <i>gestion déchet</i> échantillonnage	<b>Descripteurs (7)</b> <b>Principaux (4)</b> technologie élimination répercussions environnementales enfouissement sanitaire  <b>Secondaires (2)</b> qualité air bruit	<b>Descripteurs (18)</b> <b>Principaux (5)</b> dépotoir élimination gaz technologie étude impact <b>Secondaires (12)</b> enfouissement sanitaire incinération qualité air analyse chimique échantillonnage effet serre odeur bruit circulation avion législation règlement <b>Descripteurs Géo. (1)</b> Montréal
<b>Descripteurs Géo. (3)</b> carrière Miron CUM Montréal <b>Candidats descripteurs (4)</b> biogaz ISCST logiciel torchère	<b>Descripteurs Géo. (1)</b> ville Montréal  <b>Candidats descripteurs (1)</b> caractérisation	<b>Descripteurs Géo. (1)</b> Montréal  <b>Candidats descripteurs (1)</b> biogaz	<b>Candidats descripteurs (4)</b> biogaz cogénération méthane torchère flamme visible

Feuilles brouillon et listes de descripteurs pour le document :  
*Compostage des boues résiduaires urbaines.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Feuille brouillon</b> boues résiduaires compostage agricole boues liquides traitement des eaux usées traitement (lié aux différentes boues) des boues (lié à compostage) France boues d'épuration procédés, compostages (lié France) épandage filtration commerce facteurs économiques marché coût station d'épuration	<b>Feuille brouillon</b> compost boues ville communauté urbaine bassin valorisation déchets agriculture technologie traitement boue France	<b>Feuille brouillon</b> compostage boue valorisation traitement eau usée agricole (lié à valorisation) agriculture (lié à valorisation) biomasse technologie France station épuration aspect économique coût liquide (lié à boue) déshydratée (lié à boue)	<b>Feuille brouillon</b> boue valorisation compostage épuration (lié à boue) technologie agricole (lié à valorisation) bilan guide station d'épuration coût recyclage humus amélioration sol France
<b>Tous les champs (6)</b> <b>Descripteurs (6)</b> <b>Principaux (5)</b> épuration des eaux usées, système d' déchets liquides composts et engrais naturels agricoles, déchets coûts/bénéfices, analyse <b>Secondaires</b>	<b>Tous les champs (6)</b> <b>Descripteurs (6)</b> <b>Principaux (3)</b> composts et engrais naturels zones urbaines système d'épuration des eaux usées <b>Secondaires (2)</b> techniques de conversion des déchets économique, développement	<b>Tous les champs (7)</b> <b>Descripteurs (6)</b> <b>Principaux (5)</b> déchets liquides composts et engrais naturels zone urbaine système d'épuration des eaux usées biogradation <b>Secondaires</b>	<b>Tous les champs (8)</b> <b>Descripteurs (8)</b> <b>Principaux (6)</b> usine de traitement des eaux d'égout technique de conversion des déchets récupération des déchets composts et engrais naturels amélioration du sol recyclage <b>Secondaires (2)</b> terres agricoles déchets liquides
<b>Descripteurs Géo. (1)</b> France <b>Candidats descripteurs</b>	<b>Descripteurs Géo. (1)</b> France <b>Candidats descripteurs</b>	<b>Descripteurs Géo. (1)</b> France <b>Candidats descripteurs (1)</b> compostage	<b>Descripteurs Géo.</b> <b>Candidats descripteurs</b>

Feuilles brouillon et listes de descripteurs pour le document :  
*Étude de la possibilité d'utiliser les énergies renouvelables (éolienne, solaire, et de la biomasse) au village de Ndia Goureye au Sénégal.*

EXPERT 1	EXPERT 2	EXPERT 3	EXPERT 4
<b>Feuille brouillon</b> Sénégal Ndia Goureye énergie renouvelable éolienne solaire biogaz Sahel pays en voie de développement énergie (lié à éolienne et solaire) thèse vent radiation méthane alternative (lié à énergie renouvelable) énergie en eau cuisson électricité (lié à énergie) recommandations biomasse (lié à biogaz) énergie douce infrastructure cuisine	<b>Feuille brouillon</b> énergie renouvelable (lié à énergie) Ndia Goureye Sénégal agriculture climat technologie pays en voie de développement région rurale	<b>Feuille brouillon</b> énergie renouvelable éolienne solaire biomasse Sénégal biogaz production électricité pays en voie de développement équipement exhaure de l'eau puits (lié à éolienne) cuisson (lié à solaire) méthane gaz carbonique fumier matière organique éclairage (lié à biogaz)	<b>Feuille brouillon</b> Sénégal Sahel climat pluie saison vent sol ensoleillement désertification végétation agriculture monoculture exhaure de l'eau cuisson aliment éclairage éolienne énergie solaire recyclage déchet animal biogaz électricité (production) eau puits
<b>Tous les champs (9)</b>	<b>Tous les champs (12)</b>	<b>Tous les champs (6)</b>	<b>Tous les champs (22)</b>
<b>Descripteurs (8)</b> <b>Principaux (5)</b> énergie solaire énergie de la biomasse énergie éolienne pays en développement ressources renouvelables	<b>Descripteurs (12)</b> <b>Principaux (5)</b> énergie éolienne énergie solaire énergie de la biomasse pays en développement utilisation des ressources locales	<b>Descripteurs (6)</b> <b>Principaux (4)</b> énergie solaire énergie éolienne bouse comme combustible pays en développement	<b>Descripteurs (22)</b> <b>Principaux (12)</b> climatologie désertification zones arides sources d'énergie non-polluante ressources renouvelables énergie solaire énergie éolienne bouse comme combustible puits <i>approvisionnement en eau</i> zones rurales énergie électrique <b>Secondaires (8)</b> précipitations vents changement climatique sécheresse utilisation des ressources locales éclairage biogaz facteurs socio-économiques <b>Descripteurs Géo. (2)</b> Sénégal Sahel
<b>Secondaires</b>	<b>Secondaires (5)</b> approvisionnement en eau dans les zones rurales éclairage biogaz <i>fumier</i> sources d'énergies	<b>Secondaires (1)</b> source d'énergie non-polluante	
<b>Descripteurs Géo. (3)</b> Sénégal Ndia Goureye Sahel	<b>Descripteurs Géo. (2)</b> Ndia Goureye Sénégal	<b>Descripteurs Géo. (1)</b> Sénégal	
<b>Candidats descripteurs (1)</b> rapport de recherche	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>	<b>Candidats descripteurs</b>

## Annexe C

Protocoles concomitant et consécutif  
de l'indexation du document *Étang aéré* par E3

Indexation du document *Étang aéré* par E3 :  
Protocoles concomitant et consécutif

---

(Lecture de la page couverture) Société Québécoise d'assainissement des Eaux.  
Programme d'assainissement des Eaux du Québec (P.A.E.Q.). St-Sauveur-des-Monts, St-  
Sauveur Paroisse,

expé: Quand t'as lu ça, est-ce qu'il y a quelque chose en lisant la page titre?

E3 L'auteur.

expé: T'as identifié l'auteur? Mais à ce moment là, quand tu l'as fait. J'aimerais que tu me répondes selon ce que tu pensais au moment où tu l'as fait et non rétrospectivement.

E3 Là en fait, je lisais. Je ne pensais pas à rien de spécial, comme ça se présentait.

Piedmont. Étude de sites d'implantation d'étangs aérés. *Étangs aérés* là, je vais écrire ça.  
**Étangs aérés.**

expé: Et tu l'écris tout de suite?

E3 Oui, parce que je sais qu'il est dans le thésaurus. Donc je l'ai écrit tout de suite pour ça. Parce que j'ai écrit aussitôt que je le lis, les mots qui me semblent qui pourraient être des descripteurs je les écris tout de suite en partant. Donc j'écris tout ce que je pense d'un premier jet et après j'y vais un peu par élimination. Surtout *étang aéré*, je vois que c'est vraiment dans le titre, et je sais qu'en plus on l'a dans le thésaurus.

Consultants. Société Québécoise d'assainissement des eaux. Étangs aérés à Piedmont.  
Piedmont.

expé: Là tu sépares ta feuille. Juste "for the record", comment tu la divises?

E3 Descripteurs principaux, secondaires, candidats ou géo là.

expé: Quand t'as lu la page titre, tu mets étangs aérés. Qu'est-ce qui fait que tu mets celui-là et pas les autres? en premier comme ça?

E3 Parce que *étude de site, étang aéré* en le regardant, j'imagine que c'est le sujet principal. *Étude de site d'implantation*, ça peut être toutes sortes d'aspects qui ont rapport avec les *sites d'implantation*. En regardant à l'intérieur, je vais voir que c'est *l'évaluation des sites*. Mais ça pourrait être le choix de différents sites s'ils n'étaient pas choisis ou.

*Étang aéré*, je sais, tout de suite je fais un lien avec le terme qu'il y a dans le thésaurus. Je suis plus porté à l'écrire peut-être que *site d'implantation* qui n'est pas comme tel dans le thésaurus. Peut-être pour ça, involontairement *étude de site d'implantation d'étang aéré*, ben je sais que *étang aéré*, c'est sûre que le document il va parler de ça.

expé: Es-tu sûre, dès ce moment là tu le marques tout de suite? As-tu ton idée qui va se ramasser ici (liste des descripteurs)?

E3 Non pas nécessairement. Dans un premier temps, je sors plutôt qu'est-ce qui pourrait être là. Mais par après je vais analyser un peu et je vais voir si je le laisse là ou non. Parce que dans le fond, le titre pourrait peut-être ne pas refléter vraiment le contenu. Je pourrais

être amenée à le changer en regardant. Mais au premier abord, je trouve que ça l'air clair que ça va parler *d'étang aéré*. Ça risque d'être là (liste de descripteurs) mais ça peut changer.

Tiens, **Piedmont!**...

expé: Tu l'as écrit après avoir lu cette page titre là (*la deuxième*)?

E3 Oui, *Piedmont*, je l'ai écrit parce qu'il était là. Je suis habituée de mettre des descripteurs géographiques, tout de suite j'ai identifié que c'était à *Piedmont*. Je l'ai écrit pour ça.

expé: Plus *Piedmont* que *St-Sauveur-des-Monts* ou *St-Sauveur-Paroisse*?

E3 Oui, parce qu'il était vraiment dans le titre. Mais par après je les ai écrits quand même. Mais là, ça faisait vraiment partie du titre *étang aéré à Piedmont*, donc j'ai mis tout de suite *Piedmont*.

Étangs aérés, site d'implantation et critères de design. D'implantation, design des étangs. Étangs aérés, je l'ai!

expé: Qu'est-ce que t'as fait là? Design des étangs, étang aéré je l'ai. Peux-tu m'expliquer un peu qu'est-ce que tu fais?

E3 Là je me suis dit, *critères de design*, on fait le *design* de quoi? Des *étangs aérés*. Donc le sujet c'est encore, ça confirme mon choix d'avoir mis *étang aéré*. Mais je ne mettrais pas nécessairement *critères de design* parce c'est juste un aspect qu'on étudie. À moins que le document porte entièrement sur le *critère de design*. Mais là je vois que c'est dans la table des matières, critères de design des étangs aérés, ça me confirme qu'on va vraiment parler *d'étangs aérés*. Donc *étang aéré*, je l'ai là.

Évaluation des sites. Site, mais je pense que dans le descripteur, c'est *emplacement*. Dans le thésaurus, je pense que c'est *emplacement*. (Vérifie dans le thésaurus) Emplacement.

expé: Tu voulais marquer site?

E3 Oui. Là je vois *évaluation des sites*, une grande partie du document c'est les *sites*. Je me suis dit, on va évaluer les sites. C'est presque tout le document, la conclusion est là (*en regardant dans la table des matières*). Puis *site*, je sais que nous autres on ne l'a pas employé. Ça déjà porté à discussion, est-ce qu'on emploie *site* ou *emplacement*. Je savais tout de suite que c'est *emplacement* qu'on avait.

expé: T'es allée vérifier tout de suite dans le thésaurus?

E3 Oui. Je voulais voir. Tant qu'à en écrire un, aussi bien mettre tout de suite celui qui est dans le thésaurus. Et j'étais pas mal sûre que c'était *emplacement* et de toute façon à 99 %.

emplacement... Em, emplacement, emplacement, c'est ça. **Emplacement**. Puis *étangs aérés*, je vais le vérifier tout de suite. Je pense que c'est...

expé: Tu vérifies étang aéré?

E3 Je l'avais mis *étang aéré*, j'étais sûre aussi à 99 % qu'il était dans le thésaurus, il reste toujours le petit 1 %, mais tant qu'à le mettre; je savais que le document allait parler de *l'emplacement des étangs aérés* et j'étais sûre que c'est ça que j'allais garder dans les descripteurs principaux. C'est pour ça que tout de suite je les ai vérifiés. Ça va être fait, j'aurai pas besoin de les vérifier tantôt quand je vais les réinscrire au propre.

expé: Tu saisis rapidement. Ça fait deux minutes à peu près et t'as déjà une idée? Ça se saisit vite comme ça? y a-t-il un truc?

E3 Oui. oui. Ça dépend des documents. Celui-là, tu vois, *étangs aérés, évaluation des sites pour les étangs..* oui, oui, je savais tout de suite que le document portait sur des *sites d'étangs aérés*. L'*évaluation* ça va venir un peu plus tard je pense?

Puis souvent j'essaie de me faire une phrase: Le document porte sur l'*évaluation des emplacements des étangs aérés...* j'essaie de me faire une phrase qui résume un peu le sens du document. Souvent c'est ça dans ma tête.

expé: Est-ce que tu fais des résumés?

E3 À l'occasion. Mais il me semble qu'on avait appris cela à l'EBSI dans notre cours d'indexation. Me semble que moi à un moment. Je n'avais pas pris la spécialisation indexation, mais je me rappelle d'avoir fait cela en faisant des travaux en équipes : "ce document là parle de cet aspect là, de ce sujet principal, ou...".

expé: Et toi t'as gardé cette façon...

E3 Oui.

expé: Y a-t-il d'autres trucs?

E3 Non. Mais à un moment j'avais fait venir la norme ISO.

expé: Comment t'en ai venu à être indexeur?

E3 Parce que c'est cette job qui s'est offerte. J'ai étudié en bibliothéconomie à l'Université de Montréal avec la spécialisation informatique documentaire. Quand j'ai commencé ici, ils avaient besoin de quelqu'un responsable du traitement. Ça aurait pu être autre chose, mais je suis tombée là-dessus. Mais j'avais pas plus d'expérience qui faut là dedans.

expé: T'as suivi juste le cours qui était obligatoire en indexation là-bas?

E3 Oui. C'est sûr que les premiers que j'ai fait quand je les regarde aujourd'hui, c'est sûr que je ne ferais pas la même chose. Même, vu que j'ai travaillé sur le thésaurus, je me suis établi des règles pour tel type de document. Si c'est un rapport, si c'est une étude d'impact, répercussions environnementales, on se l'est dit, on le met toujours en descripteur principal ou si c'est un rapport d'audience, les rapports du BAP, les rapports d'enquête et audience publique. Parce qu'ils ont toujours un aspect sur les répercussions environnementales, là on le met en DE2. Le but c'est une audience donc on met audience publique en DE1 et DE2 répercussions environnementales parce que c'est une partie du document. Ce n'est pas tout le document. Ça c'est des règles qu'on s'est établi à force de le travailler (*le thésaurus*) et d'en discuter ensemble, avec l'autre indexeur, E2, même celle qui fait la référence. Parce que quand je suis arrivée, il y a beaucoup de termes qui se sont ajoutés, comme développement durable, ou des termes ben à la mode. Pour en arriver à améliorer le thésaurus, on a fait des rencontres. On prenait des notes en indexant

« tiens ce terme là, ça nous manque ». Celle qui fait la recherche sait comment les documents sont demandés. Pendant 2 ou 3 mois, on faisait une rencontre toute les semaines. On a tout passé le thésaurus. On se disait là j'ai telle telle note, est-ce qu'on l'ajoute dans le thésaurus ce mot? Pourquoi je raconte tout cela?

Après avoir fait cet exercice de la mise à jour du thésaurus, ça facilite mon ouvrage.

expé: Comment a été formé ce thésaurus?

E3 La banque Envirodoq date de la fin '70 début '80, elle était interrogeable sur un système IST informatec. Mais ça n'existe plus ce serveur. C'était une bibliographie sur l'Environnement au Québec. Gérard pourrait plus t'en raconter sur cet historique là, moi je n'étais pas là.

Des fois, j'ai déjà vu comme auteur (*de la notice*) une autre personne qui a travaillé ici, du thésaurus auteur. J'imagine qu'elle a du travailler sur le thésaurus de départ. Avant, les documents que l'autre indexeur traite étaient traités avec les vedettes-matières. Avant que j'arrive la décision a été prise qu'il n'y aurait plus de vedettes-matières. L'autre indexeur se servirait du thésaurus. Ça fait trois ans que je travaille ici.

expé: Quand tu fais l'indexation, penses-tu à d'autres documents qui ont été indexés avec ces mêmes termes? Si on regarde cette indexation, est-ce que t'as une idée que t'as d'autres types de documents qui ressemble à lui au niveau de leur description?

E3 Oui.

expé: Est-ce que ça t'aide cela pour choisir des termes, de te dire que ce document ressemble à ce type?

E3 Pas celui-là comme tel. Mais pour certains documents, il arrive que je retourne dans des anciens documents. Et avec l'ordinateur, ça va bien, juste à pitonner. Surtout comme des séries. Des fois je me dis, comment ça qu'on a mis ça dans ce temps là?

c'est ça aussi. *Étangs aérés, emplacement. Évaluation des sites, contraintes d'évaluation des sites, contraintes environnementales, contraintes physiques, contraintes techniques, contraintes monétaires...*

expé: Fais-tu quelque chose là, penses-tu à qu. ch.?

E3 Non. Je ne pense pas.

Grille d'évaluation des sites. Évaluation détaillée des sites. Site usine actuelle, variante - étangs aérés. Liste des figures, liste des tableaux. Introduction. Les municipalités de Piedmont, St-Sauveur-des-Monts et St-Sauveur-Paroisse se sont regroupées dans le but d'effectuer un traitement commun de leurs eaux usées. Traitement eaux usées.

expé: Tout de suite comme ça?

E3 Oui. Si on évalue les *emplacements* pour des *étangs aérés*, ça veut dire qu'on a l'idée, c'est pour traiter des *eaux usées*. Le but c'est ça. De traiter des eaux usées. Donc là je pensais, si quelqu'un fait une recherche un peu plus générale sur le *traitement des eaux usées*, il va être intéressé à avoir ce document là. C'est pour cela que je l'ai retenu. Mais je ne savais pas encore si je le mettais en descripteur principal ou secondaire. Je l'ai marqué comme cela.

expé: Ceux là, tu le savais à un moment qu'ils étaient pour se ramasser là?

**E3** Oui. *Emplacement et étang aéré*, je savais que ç'allait être dans les descripteurs principaux. En plus, ils étaient dans le titre. Vraiment, le document il parle de l'*évaluation des sites* donc...

expé: Puis quand t'as regardé, t'as passé rapidement sur la liste des figures, après t'arrives à l'introduction, tu t'y arrêtes. Est-ce que c'est utile pour toi la liste des figures ou...

**E3** Pas dans un premier temps. Pas plus qu'il faut. Peut-être pour compléter ou réviser en dernier. Je regarde s'il y a des choses importantes dans la liste des figures ou des tableaux, et dans les annexes et si il y a quelque chose qui mériterait vraiment la peine d'être mentionné ou d'être souligné, ou sinon, pas dans un premier temps, c'est rare.

expé: Quand tu dis, si je reviens pour réviser ou des trucs importants, qu'est-ce qu'un truc important?

**E3** S'il y avait une annexe qui prend, le tiers du document, des relevés d'échantillons de la qualité de l'air avec toute sorte de substances puis des échantillons qui sont pris. Là peut-être que je le marquerais en descripteur principal... *échantillonnage* ou quelque chose du genre. Mais vraiment si ça un, j'y vais peut-être par volume, le nombre de pages.

Traitement eaux usées. Là, *traitement eaux usées*, je ne le vérifie pas parce que je sais qu'il est dans le thésaurus. Là, je ne sais pas si j'ai besoin de te le dire comme ça, mais dans ma tête je le pense. (Expé: Oui, c'est tu n'as qu'à le dire) OK. Traitement eaux usées. (Expé: mais dis, comme tu le penses dans ta tête et non de me justifier à moi. C'est ça, tu sors ce que tu as dans la tête.) OK. Dans le but d'effectuer un traitement commun de leurs eaux usées se déversant dans la Rivière du Nord. Piedmont!... À St-Sauveur-des-Monts, je vais rajouter ça. Puis **St-Sauveur-Paroisse**. *Paroisse*.

expé: Tu reviens, tu dis *Piedmont*, tu regardes sur la page titre et là tu ajoutes...

**E3** Là j'ai rajouté *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*. J'ai pensé, *Rivière du Nord* c'est un lieu cela, je vais sauter dans les lieux là. *Rivière du Nord* ça m'est passé par la tête, je le mets-tu ou je ne le mets pas. J'ai attendu. Mais ici on parlait de la municipalité de *Piedmont*, *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*. Alors je les ai écrits pour pas les oublier, vu qu'on était dans les descripteurs géographiques.

expé: Qu'est-ce qui est arrivé avec Rivière du Nord? Tu t'es comme arrêté dessus...

**E3** J'ai attendu, je ne le sais pas pourquoi.

expé: Tu ne sais pas pourquoi, mais par contre ceux-là tu les as marqués. Pourquoi?

**E3** Parce que peut-être aussi inconsciemment j'ai pensé aussi, on fait l'étude de *sites d'implantation d'étangs aérés*. Si on fait l'étude de *sites d'implantation*, le fait que les *eaux usées* se jettent dans la *Rivière du Nord* ça plus ou moins d'importance. Mais si les *sites* peuvent se retrouver à *Piedmont*, *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*. Donc s'il y en a là, ça peut être important, pis *Piedmont* aussi. J'ai dû rajouter *St-Sauveur-des-Monts*, *St-Sauveur-Paroisse*, pis *Rivière du Nord*, étant donné qu'on n'en parle pas. Que les *eaux usées* se déversent dans la rivière du Nord ou dans une autre, ce n'est pas le but, ce n'est pas le document. Je me suis dit que ce n'était pas le document.

Traitement commun de leurs eaux usées se déversant dans la Rivière du Nord. Une épuration par biofiltration ou par étangs facultatifs aérés leur a été proposée...

expé: As-tu fait qu. ch. avec ça là?

E3 Non. Je ne m'en rappelle pas.

L'emplacement déterminé pour les étangs aérés étant situé dans une zone à vocation résidentielle, a mandaté le consortium Gendron afin d'identifier des sites alternatifs d'implantation de l'usine d'épuration. Ces sites, parmi 5 différents emplacements potentiels situés...

expé: Quand t'as dit usine, t'as regardé ta feuille pis t'es revenue? Te souviens-tu?

E3 Non.

Usine d'épuration, ça, je ne me rappelle plus jamais, si c'est usine ou usine... station. Je pense que c'est station.

expé: Qu'est-ce qui se passe là?

E3: On parlait d'usine pour l'implantation d'usine d'épuration. Là je savais que ça parlait d'étang aéré mais peut-être que ça pouvait allé plus loin aussi. Ça pouvait aller jusqu'à l'usine, construction de l'usine ou.. J'ai pensé l'écrire usine, comme ça pour pas l'oublier si jamais le document va plus loin là-dedans. Mais usine d'épuration aussi a fait l'objet de discussion parce souvent on cherchait usine d'épuration pis ce n'est pas ça qu'on emploie. C'est station d'épuration. C'est peut-être pour cela que je me pose la question là.

expé: Là tu vas vérifier directement dans le thésaurus?

E3 Oui.

expé: Pourquoi plus station d'épuration que usine d'épuration? Le terme technique le plus?

E3 Moi non plus. Je ne sais pas. Parce que dans des documents des fois, on va parler de station d'épuration, des fois d'usine d'épuration. Mais je me demande si c'est dans le nouvelle version, je ne sais pas si je l'ai fait, mais il fallait que je crée un renvoi: usine d'épuration employer station d'épuration. Parce qu'il faut qu'on en choisisse un. Puis vu que c'est tout le temps station d'épuration, il était là quand je suis arrivée. Tant qu'à le changer, autant créer un renvoi. Parce que le renvoi est utile, je me rappelle que l'autre indexeur est déjà venu dans mon bureau « quand on parle d'usine d'épuration, on ne sait jamais si ...il n'est pas dans le thésaurus, on ne sait pas qu'on doit prendre station ».

expé: Est-ce qu'il y en a beaucoup de cas comme ça?

E3 Oui. Des renvois c'est ce que j'ai le plus à rajouter.

expé: C'est un vocabulaire qui est riche?

E3 Oui. Parce qu'il y a beaucoup de termes dans le thésaurus. Je ne sais pas combien.

Station... (vérifie dans le thésaurus) Station d'épuration pour traiter les eaux usées... Un autre station de purification. Ça, c'est pour l'eau potable, alors c'est station d'épuration. Station d'épuration.

expé: T'as pas juste regardé *station d'épuration*, t'as regardé *station de purification*, puis un autre petit truc en bas, l'eau...

E3 (on regarde ensemble dans le thésaurus). Là je cherchais *station de purification* parce que je savais qu'il y en avait un pour l'eau potable et l'eau usée. *Station de purification* pour l'eau potable. Pis j'ai vu *station d'épuration* pour les eaux usées. Mais vu que dans ma tête, je me disais, bon on a *station de purification* pour l'eau potable, je suis allé vérifier si il était bien là. Mais ça n'avait plus rapport avec mon travail. C'était une curiosité personnelle, pour vérifier que ma pensée était correcte.

expé: Ça te confirme ta connaissance du thésaurus.

E3 Oui, c'est un peu ça. Puis usine qu'est-ce qu'on a mis (regarde dans le thésaurus). Ah oui! *usine*, le renvoi est là. Mais *usine de purification* n'est pas là. C'est là le problème.

Cinq différents... Bon! Chaque usine sera jugée dans le présent document en fonction de contraintes environnementales. Le présent rapport résume les avantages et les inconvénients de chacun.

expé: Tu passes rapidement quand tu lis les *contraintes*. T'as rien traité au niveau des contraintes? C'était important ou non pour toi vu que t'as passé vite?

E3 Non. Qu'est-ce qui est important c'est qu'on les évalue. Mais les types de contraintes, non je n'avais pas jugé plus que ça.

expé: Vu qu'elles occupent une bonne partie de la TM, si tu faisais un rapport entre les deux?

E3 Puis *contraintes*, je sais qu'il n'y a rien dans le thésaurus qui va réussir à rendre cela. Donc ça règle mon problème. Si ça avait été vraiment un sujet d'importance, je l'aurais probablement mis dans les candidats. *Contraintes* je savais que de toute façon je ne pouvais pas le rendre.

expé: Ta connaissance du thésaurus ça t'aide à être plus efficace dans ta job?

E3 Ben oui, sûrement. Ça va paraître peut-être avec l'autre (*Infoterra*)?

... monétaire des terrains ainsi que des dommages potentiels à défrayer pour l'acquisition de sites. Étangs aérés, sites d'implantation et critères de design. Sites d'implantation. (Regarde les termes inscrits sur sa feuille et fait un crochet près du terme emplacement) Emplacement, je l'avais...

expé: Quand tu lis tu regardes ta feuille et tu fais un petit crochet, qu'est-ce qui se passe?

E3 Ça m'a confirmé qu'*emplacement* c'était bon. C'était pour me dire, celui-là je le garde.

Les sites d'implantation à Piedmont sont connus sous les appellations suivantes: usine, Lutfy. Qu'est-ce que c'est que ça?

expé: Là tu trouvais cela? C'est pas des trucs que tu marquerais?

E3 *Lutfy*. Non. C'est des *sites* parmi tant d'autres. À moins que, ça dépend du niveau d'indexation que tu fais. Une indexation ben ben détaillée. Ben non, de toute façon, je n'avais pas le goût de mettre cela.

E3 Avoir 5, 6 descripteurs principaux c'est bon. Descripteurs secondaires (DE2) c'est plus les aspects par chapitre ou quand c'est disons... Moi je suis portée à mettre là des termes plus généraux. Ce n'est pas le sujet précis, en DE2. Mais on n'a pas vraiment de politiques écrites là-dessus. Ça fait partie de mon travail aussi. Je n'ai pas le temps. Mais ça on s'en parle un peu entre nous pour essayer d'être un peu uniforme, mais c'est sûr qu'on indexe probablement pas...

expé: Comme ici (*en DE2*) t'as mis *traitement eau usée*. C'est plus général, ce n'est pas le coeur du document?

E3 C'est pour ça que je suis portée à le mettre en ..., mais je ne sais pas, l'autre indexeur ou E2 l'aurait peut-être mis en descripteur principaux. Mais comme l'autre indexeur, lui il n'en met pas de descripteurs principaux. Il va moins dans le détail, pour la collection générale.

expé: Pourquoi? Parce que ça s'adresse à un autre public?

E3 Parce que c'est une collection qui est plus générale.

expé: Ce n'est pas des rapport comme cela?

E3 Non. Puis *Envirodoq* c'est LA bibliographie sur l'environnement au Québec. Puis l'environnement c'est à la mode, on essaie de donner le plus de chance possible à tout le monde de retrouver ce qu'il cherche, donc on y va un peu plus précis. Puis elle est en CD-ROM aussi la banque *Envirodoq*, qui est vendue dans les Universités. La collection générale c'est plus à l'interne. Peut-être aussi le fait que c'était les vedettes-matières l'idée de départ. Quoique des fois on y dit à l'autre indexeur d'aller plus (*en détails*) ...lorsqu'il y a des chapitres de plus d'importance. On commence à y en parler.

Méandre. Les sites... les cinq sites d'implantation à Piedmont, connus sous les appellations suivantes... Usine actuelle, Lutfy, Méandre, Boysfarm, Ferme Mofette... Ferme Mofette, tiens! La figure 2.1 présente leur emplacement respectif. Design des étangs. Étangs, je l'ai! Étangs.

E3 Oui. *Design des étangs*, on parle d'*étang*. Si une personne est intéressée à avoir des documents sur les *étangs* le *design*, elle va chercher avec *étang*. Ça me surprendrait qu'elle cherche avec *design*. De toute façon *design* n'est pas dans notre thésaurus. Puis il n'y a rien qui peut rendre ce terme là dans le thésaurus. Alors, je pense que si quelqu'un se présente pour avoir des modèles d'étangs, elle va chercher directement avec *étang aéré*. Elle va avoir ça. Ça va être bon.

expé: Avais-tu considéré à un moment de le prendre?

E3 Non je ne l'ai pas considéré. Pis *design des étangs*, ça m'a fait penser à *étangs*.

Design... coliformes fécaux, ça c'est sur les étangs...

E3 Je suis tombée là-dessus, j'ai dû regarder rapidement.

expé: Et t'associes cela aux étangs?

**E3** Oui.

expé: Là tu feuilletes le document, est-ce dans tes habitudes?

**E3** Oui, je fais souvent cela. Pour être sûre que je n'oublie rien.

expé: Ce que je trouve épatant, c'est que tu regardes un document et tu vas retirer ce qui est important. Mais comment en regardant le document peux-tu me dire ce qui est important? Tu peux saisir ce qui est important? Qu'est-ce que tu regardes, comment tu le regardes pour sortir tu le regardes? Parce que tu ne lis pas tout...

**E3** Je regarde les grands titres, les caractères différents les caractères gras. Je regarde vite, vite.

expé: Vite, et s'il y a quelque chose d'intéressant tu le retires. Si t'es capable de voir comment ça se passe dans ta tête pendant que tu fais cela, exemple *coliforme fécaux*.

**E3** Ben là *design des étangs*, mes yeux ont été portés plus vers le centre qu'au début. Ici je voyais toute sorte de petits... mètres cubes, peut-être que rapidement je me suis dit, c'est dans le *design* cela c'est des calculs pour des grandeurs ou des volumes. *Coliformes fécaux*, encore dans les *étangs* là, ça doit être des...je ne sais pas.

Évaluation des sites, emplacement, je l'ai...

**E3** Je vais tu y penser à *évaluation!*

Contraintes d'évaluation des sites, contraintes environnementales... Zonage actuel, contraintes...

expé: Pensaistu à quelque chose de particulier? Te souviens-tu? Tu mentionnes deux fois contraintes.

**E3** Je ne le sais pas. J'ai peut-être pensé à *contraintes... impact environnemental, répercussions environnementales*. Après ça je me suis dit, *contraintes* ce n'est pas nécessairement les *répercussions*. Pis *contraintes*, il n'y a vraiment rien qui le rend dans notre thésaurus donc j'ai passé je suis allé voir plus loin.

expé: te souviens-tu si c'est ce que tu as pensé là ou t'es en train de me dire ce que peut-être t'as pu penser?

**E3** Non je pense que j'ai pensé ça parce que je me rappelle avoir pensé un moment donné à *répercussions environnementales*. Je ne me rappelle pas si c'est vraiment ici, mais je sais que j'y ai pensé. Mais ça doit être ici. Pis je me suis dit *contraintes* pis *répercussions* ce n'est pas la même chose.

expé: Le nombre de fois que tu vois un terme est-ce que ça t'influence? Est-ce que ça te dit s'il est meilleur ou non?

**E3** Oui, mais les termes qui reviennent, contraintes c'est sûr que ça revient là parce que c'était dans la table des matières, mais ceux qui reviennent, étangs eaux usées, sites, c'est ceux qui reviennent le plus souvent. Cinq sites étudiés.

Zonage actuel, visibilité, contaminations potentielles de puits... Contraintes physiques, contraintes techniques, contraintes monétaires...

**E3** là je sautais vite par-dessus. Je voulais aller voir ce qu'il y avait d'autre.

*Usine... site... étangs aérés facultatifs... Site Lufty, emplacements, ça je les ai. Ça, c'est les emplacements...*

**E3** Là je regardais vraiment ce qu'il y avait d'autre. Tu vois ça c'est les *sites* là.

expé: Qu'est-ce que ça te disait? Il n'y avait rien qui t'accrochais plus que d'autre chose?

**E3** Non. Là je voyais qu'on évaluait chacun des *sites*. Un après l'autre. Et ça je n'avais pas le goût de m'attarder là-dessus. Non, quand même que je vais dans le détail de chacun, je trouve que ça ne donne rien. Non.

expé: C'est pour cela que tu passais vite dessus?

**E3** Oui. Puis je sais que d'un *site* à l'autre ça va tout le temps être la même chose qui va revenir. On va essayer d'évaluer un peu les mêmes choses pour chacun. En voir un ou en voir cinq ça revient au même. Je n'ai pas perdu de temps à regarder chacun en détail.

expé: Ça doit être l'habitude aussi de voir ce type de document là, que t'es capable d'identifier et de dire que cette partie là c'est pas...

**E3** Oui sûrement.

Conclusion. Bon. Cette étude a permis d'analyser

expé: La regardes-tu normalement la conclusion?

**E3** Oui. Je la regarde. Je regarde l'intro, la table des matières, le résumé, la conclusion. J'aime bien regarder la conclusion, le titre. Il y a des documents que je ne feuillète même pas. Je regarde l'introduction, la conclusion, quand c'est des évaluations environnementales des fois, c'est tout le temps le même modèle de document. C'est tout le temps la même table des matières, donc des fois je ne feuillète pas nécessairement. Surtout quand c'est des grosses briques, je ne feuillète pas. Dépendamment de mon humeur aussi. Il y a des journées qu'on est plus pointilleux, ou bien des journées on a le goût d'aller plus dans le détail, pis d'autres fois on n'est moins méticuleux.

expé: Dans les parties que tu regardes, est-ce qu'il y a des parties auxquelles tu peux dire: Dans cette partie là, je regarde tel type d'information. Je vais essayer d'aller chercher un peu comme si chacune des parties aurait son rôle ou sa fonction?

**E3** Dans l'intro, je regarde si ça résume un peu. Souvent dans l'introduction, on dit ce qu'il va y avoir dans le document. Des fois on dit dans telle partie il va y avoir telle chose, dans telle autre telle chose. Donc l'intro, c'est comme si ça reprenait la table des matières plus en langage... (fait un mouvement de continuité)

expé: Suivi.

**E3** Oui. Ça je regarde ça dans l'intro., le contenu. Parce que j'ai déjà appris cela comment faire une introduction pis la dernière partie de l'introduction c'est supposé être, décrire ce qu'il y a dans le document. Souvent il y en a qui le font. Donc dans l'intro c'est ce que je m'attends à retrouver. Pis dans la conclusion, on reprend un peu les éléments du document, normalement. Ça aussi ça m'aide.

expé: Tu regardes la conclusion pour retrouver des informations importantes?

**E3** Oui, que j'aurais pu oublier ou pour me confirmer dans mes choix.

expé: L'intro est-ce la même chose?

**E3** Là l'intro je commence dans l'intro. Je commence avec le titre, ça me donne déjà des bons points de départ. J'ai déjà *étang aéré, site*. La table des matières aussi, je vois en gros ce qu'il va y avoir. Pis là l'intro c'est un peu comme un titre, mais en plus détaillé avec vraiment le contenu du document. Et la conclusion c'est plus pour vérifier, pour voir si je n'ai rien oublié, et pour me confirmer dans mes choix.

cinq sites différents pour l'implantation des étangs aérés et... (Elle réécrit par dessus le crochet déjà inscrit près de ces termes)Emplacement, étangs aérés...

expé: Tu fais tout de suite un petit crochet à côté.

**E3** Oui je le refais.

usine de biofiltration dans la municipalité de Piedmont. Piedmont, j'ai ça. St-Sauveur-des-Monts, St-Sauveur Paroisse... Pourquoi St-Sauveur-des-Monts?...

expé: Qu'est-ce que tu te poses comme question?

**E3** Là j'avais mis *St-Sauveur-des-Monts, St-Sauveur Paroisse* parce qu'au début on parlait de...*(retourne au début du document) municipalité Piedmont, St-Sauveur-des-Monts, St-Sauveur Paroisse se sont regroupés*. Je me suis dit peut-être le document il va traiter... les sites vont se retrouver également dans une de ces trois municipalités. Ici *(dans la conclusion)*, j'ai vu dans la municipalité de *Piedmont*. Alors là j'ai le goût d'enlever *St-Sauveur-des-Monts, St-Sauveur Paroisse*. Là je suis retourné vérifier pourquoi j'avais mis ça *St-Sauveur-des-Monts, St-Sauveur Paroisse*. Voir à quelle place j'avais vu cela et si c'était justifié de le garder ou non.

expé: Et tu vas prendre la décision je crois..

**E3** De les enlever.

St-Sauveur... d'épuration par biofiltration par étangs aérés... Emplacement... dans une vocation résidentielle... afin d'identifier des sites alternatifs pour l'épuration. Les sites seront déterminés parmi 5 différents emplacements potentiels situés dans la municipalité de Piedmont. (raye les termes suivants de sa feuille) St-Sauveur-des-Monts, St-Sauveur Paroisse, je vais enlever ça.

expé: sais-tu ce que t'as décidé?

**E3** Ben les sites seront déterminés parmi 5 différents emplacements potentiels situés (elle met l'accent sur ce mot) dans la municipalité de Piedmont. Donc on voyait que les emplacements étaient dans la municipalité de *Piedmont* pas à *St-Sauveur-des-Monts, St-Sauveur-Paroisse*.

expé: Ça n'aurait pas été bon de les mettre?

**E3** Non. Ça aurait été trop. Si quelqu'un cherche des documents sur *St-Sauveur-des-Monts*, pis qu'il tombe là-dessus, peut-être qu'il ne sera pas nécessairement content d'avoir ça. Parce que on en parle pas tant que ça de *St-Sauveur-des-Monts, St-Sauveur Paroisse*.

expé: Est-ce que c'est trompeur le fait qu'on le voit sur la couverture?

**E3** Oui parce que tu vois je les ai mis au départ. Au départ j'ai pensé qu'on pouvait en discuter là-dedans.

expé: et *Société Québécoise d'Assainissement des Eaux* ça ne t'a rien sonné?

E3 Ça je le mettrais comme auteur. Parce que j'imagine que c'est une étude qui est faite par ces consultants là pour la Société. Donc ça n'aurait pas rajouté en sujet. Le document ne porte pas sur la Société, il porte sur les *étangs*, *l'évaluation*.

expé: Penses-tu vraiment en fonction de l'**usager**?

E3 Oui.

expé: Tout le temps ou...

E3 Souvent. Souvent en dernier quand je révise, je me dis si quelqu'un cherche des documents sur les *étang aérés*, il va être content d'avoir ça. *Emplacement* tout seul, ça serait surprenant que quelqu'un cherche un document juste avec *emplacement*. Donc il va chercher *emplacement* pis *dépotoir* ou *emplacement d'étangs aérés*. Pis s'il cherche *emplacement d'étangs aérés*, c'est en plein ça. Il va être ben content. *Évaluation* aussi... Là je ne l'ai pas encore ajouté *évaluation*. Mais je pense souvent à ça. Je me dis souvent, si quelqu'un cherche un document là-dessus, pis il retrouve cela, va-t-il être content? Ça des fois ça me fait enlever quand... Parce que je me dis, le document il parle de ça, mais pas tant que ça c'est minime ou... Le document parle de ça, mais indirectement. Comme tantôt, on va en avoir un exemple avec *dépotoir*. Le sujet ce n'est pas le *dépotoir*, mais je l'ai mis pareil. Mais des fois, j'aurais pu ne pas le mettre parce que ... je ne le sais pas. En tout cas, on en reparlera.

Parce que des fois, le document nous fait penser à quelque chose, nous fait penser à un sujet, indirectement ça rapport à ce sujet là. Sauf qu'on n'en parle pas vraiment dans le document. Je trouve dans ce temps là qu'on n'est peut-être pas justifié de le mettre. Parce que si une personne cherche quelque chose avec le sujet pis, il retrouve le document pis on n'en parle pas vraiment. Mais ça a un lien quelconque avec. Il ne sera peut-être pas content de l'avoir. Parce que ça fait du bruit.

expé: Est-ce que ça arrive souvent ça qu'en regardant un document, il y a d'autres trucs qui s'en viennent?

E3 Oui. Des fois, peut-être dans 10 % des cas. Ça arrive temps en temps.

expé: Et l'**usager** auquel tu penses, a-t-il...

E3 un profil?

expé: Oui. À qui tu penses quand tu penses usager?

E3 C'est vague. Je pense plus aux gens qui travaille au ministère de l'environnement, pis qu'ils ont affaire... aux chercheurs qui travaillent au ministère de l'environnement. Je pense que c'est plus ça que j'ai en tête. Parce que notre premier but c'est de desservir la clientèle ministérielle. Pis le ministre c'est rare qu'ils vont venir faire des recherches. Les étudiants ou monsieur tout le monde, ils sont les bienvenus, mais notre but premier ce n'est pas de leur répondre.

expé: Dans ce sens là, il n'aurait pas été utile de mettre, t'as mis *traitement eau usée*, est-ce qu'on aurait pu mettre *assainissement des eaux*?

E3 Oui, c'est vrai. Parce qu'on l'a *assainissement* tout court. J'aurais pu, mais je n'y ai pas pensé.

expé: Est-ce que ça aurait été trop général, face au type d'**usager** que tu regardes?

E3 Non, ç'aurait pu être bon. Tu me le fait penser, j'aurai peut-être du.

expé: Oui mais il faut que le document en parle?

E3 Oui, ça en parle de *l'assainissement des eaux*.

expé: C'est pas trop général?

E3 Mais *traitement eau usée* pis *assainissement*. Peut-être que j'aurais mis *assainissement des eaux* si on avait parlé des programmes d'assainissement des eaux du ministère ou de la politique du gouvernement en matière d'assainissement des eaux. Là j'aurais été plus portée à mettre *assainissement des eaux*, parce que c'est le nom d'un programme. Programme d'assainissement des eaux du Québec. Je me dis si les gens cherchent des documents sur les politiques ministérielles en matière *d'assainissement des eaux*, là je pense que je l'aurai mis.

expé: Est-ce comme l'exemple de *répercussions environnementales*, il y a une sorte d'entente implicite que quand il y a des documents sur des *impacts environnementaux*, on met *répercussions*?

E3 non parce qu'on n'a jamais discuté de ça. C'est personnel dans ma tête.

expé: Ok. C'est pourquoi je me demandais quel usager. Parce que monsieur tout le monde ne connaît certainement pas les étangs aérés.

C'est juste Piedmont. Chaque site sera jugé... l'analyse, les... C'est pour le traitement des eaux usées. Ça, ça va. Implantation d'étangs nécessitant dix hectares comme zone... L'étang d'implantation sur le site de l'usine actuelle nécessite... Bon! une expropriation... L'implantation d'une usine de biofiltration sur le site adjacent, telle que proposée...

expé: As-tu pensé à quelque chose quand t'as lu l'usine de biofiltration, t'arrêtes un peu là?

E3 Non.

expé: As-tu considéré *biofiltration* dans ton truc?

E3 non.

que peu de superficie. Un dézonage est cependant requis ainsi qu'une dérogation... Le site boysfarm, le site Lutfy, le site Méandre... qu'aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs aérés. **Station d'épuration**, je vais l'enlever aussi (~~raye station d'épuration~~).

expé: Pourquoi t'as enlevé station d'épuration?

E3 Là j'avais vu que dans le document, il avait une grosse partie qui concernait les *contraintes, design des étangs, évaluation des sites*... Ça avait rapport aux *étangs*. Les *sites* c'était pour *l'implantation des étangs*. Pis ici on disait, la grille d'évaluation des sites fait voir qu'aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs. Donc là vraiment, le document il portait sur *l'implantation d'étang*, il ne portait pas sur la *construction d'une usine*. C'est pour cela que je l'ai enlevé. Je me disais qu'*usine* on en parle parci parlà, mais ce n'est pas le document, c'est pas vraiment ça.

Tous les sites présentent en effet diverses con... con... de site... Ça va être étang aéré... (regarde sa feuille brouillon et le document)

expé: Tu reviens à la page titre, te souviens-tu ?

E3 C'est pas évident.

emplacement, traitement eaux usées... (*ouvre le document à la table des matières*)  
Répercussion... Répercussion, répercussion...

E3 Là je pense à *répercussions* parce que j'ai pensé *contraintes*. C'est ça que je te disais tantôt. Pis après ça j'ai pensé *contraintes et répercussions* ce n'est pas tout à fait la même chose.

expé: Te souviens-tu ce que tu faisais quand tu regardes la page titre et ..

E3 Là je me disais qu'en mettant *étang aéré* je ne me trompais certainement pas, *emplacement* et *traitement eau usée* également. Parce que le but des *étangs aérés* c'est le *traitement des eaux usées*. Avec ces termes là, n'importe qui qui cherche avec ces termes va être content d'avoir ça. Parce que tout a rapport. Ça me confirmait dans des choix de ces termes là, parce que j'étais sûre que le document traite de ça.

Contraintes d'évaluation des sites, contraintes environnementales. Grille d'évaluation, évaluation détaillée des sites. Évaluation, répercussion...

expé: T'as nommé répercussion mais tu n'as rien fait avec, te souviens-tu?

E3 Oui. Dans ma tête ça s'est passé, *contraintes* et *répercussions* ce n'est pas la même chose. On parle des *contraintes* mais pas des *répercussions* sur l'environnement de faire des *sites*. Si on en parle c'est une phrase par-ci par-là, mais je ne suis pas allé voir dans le détail parce qu'il n'y avait pas de grand thème qui parlait de *répercussions*. Alors tout de suite je l'ai abandonné.

expé: N'empêche que tu y as pensé?

E3 J'y ai pensé, je pense en voyant *contraintes environnementales*. C'est un grand mot, un grand descripteur. C'est comme un mot qui s'associe.

Évaluation. Évaluation... (*cherche dans le thésaurus*)

expé: Savais-tu si évaluation existait?

E3 Oui. J'étais pas mal sûre. Mais je ne savais pas s'il existait tout seul. Parce que des fois il y a des termes qui existent tout seul comme *évaluation* pis il peut y avoir quelque chose d'autre. Donc je suis allé voir vraiment si c'était... (*va voir dans le thésaurus*) c'est ça que je suis allé vérifier.

Évaluation des sites.

expé: Ce n'est pas un terme qui est très général?

E3 Oui, c'est un terme qui est très général. Ça c'est un terme qui est dans notre thésaurus. Dans le fond, on pourrait s'interroger sur la valeur d'avoir ce terme là dans notre thésaurus. Mais là, il est là, et vu que ça irait bien et que ça irait bien avec ce document là, je l'ai mis. Parce que dans le fond, c'est sûr que c'est un terme qui est super général. Peut-être que ça ne vaut pas la peine d'avoir mis cela dans le thésaurus...

Pis ici (*dans la table des matières*) on voit la grosse partie qui parle de *l'évaluation des sites*. Donc si quelqu'un fait une recherche vraiment poussée pis il veut avoir *l'évaluation d'emplacement* pour telle construction. Ben peut-être qu'il va le mettre. Si une personne cherche seulement avec *étang aéré*, elle va retrouver toute sorte d'affaires sur les *étangs aérés*: les *répercussions environnementales*, situé dans telle municipalité. Mais si elle trouve que c'est trop, pis elle veut spécifier, elle veut *l'évaluation* soit des *emplacements* ou des *étangs aérés* ou.. elle peut spécifier avec. Mais je ne suis pas sûre qu'une personne va chercher aussi spécifique que ça. Ça peut arriver, mais dans l'immédiat peut-être pas parce que peut-être qu'en cherchant juste avec *étang aéré*, il n'y aura pas tant de documents que ça qui vont sortir. Elle n'aura peut-être pas à fouiller tant que ça pour trouver ce qui l'intéresse. À l'avenir, on ne le sait pas par exemple. Parce que si la banque devient grosse grosse, peut-être que là ça peut être utile d'avoir ce terme là.

Usine actuelle... (parcours la liste des figures dans le document) Site d'implantation. Plan d'implantation... Schéma d'implantation...

expé: Tu regardes la liste des tableaux? Et tu nommes souvent implantation.  
E3 C'est un mot qui est central dans la phrase. C'est peut-être pour ça que je l'ai vu.

Implantation, j'ai-tu ça?... Implantation... (consulte le thésaurus)

expé: tu vas voir à implantation?  
E3 ben oui, je suis allé.  
expé: Qu'est-ce que tu fais avec?  
E3 Il n'est pas là, je ne l'ai pas mis.  
expé: mais t'es allé voir parce que tu considérais que t'aurais pu le mettre? Te souviens-tu?  
E3 S'il avait été là? Peut-être, ou j'aurais regardé les liens qui étaient avec *implantation*. Dans les figures il y avait souvent *implantation*. Il était dans quasiment chaque phrase. C'est peut-être pour cela que je l'ai dit.

implantation de... Visibilité, schéma d'implantation... Municipalité de Piedmont...

expé: et là tu reviens à l'intro encore une fois...  
E3 Je dois faire ma révision finale. Pour être sûre que je n'ai rien oublié.

Usine d'épuration. Ces sites seront déterminés parmi cinq différents emplacements... chaque site sera jugé d'après le rapport et les avantages et les inconvénients de chacun. Il fait également... Bon! (regarde sa feuille) Mon document porte sur l'évaluation des emplacements pour les étangs aérés.

expé: Qu'est-ce que tu fais là?  
E3 Je fais ma petite phrase. Mon document il parle de ça. Là ça devait tirer à sa fin. En dernier je fais une petite révision finale. Mon document il parle de ça et je me dis aussi si une *personne* vient et cherche là-dessus, il ne sera pas mécontent d'avoir ça. Ça va faire son affaire. En dernier, je reviens tout le temps là-dessus.

*Traitement eaux usées*, c'est plus général, je pourrais le mettre dans descripteurs secondaires ou principal?

expé: En même temps que tu fais ta petite phrase, tu les places?

E3 Oui.

expé: Les places-tu avant? Est-ce que ça t'arrive de les placer avant, même si tu ne les écris pas comme tel, de savoir dans ta tête où il vont aller?

E3 Oui. Parce que là tu vois je n'ai rien écrit, mais d'habitude, des fois j'en écris quand c'est un document plus volumineux, et qu'il y a toutes sortes de petits chapitres... parce je fais tout le temps mes colonnes comme ça et ici c'est tout le temps DE1 DE2 et DEgéo. Mais là je ne sais pas, ça peut-être pas adonné parce que j'avais peut-être pas des documents qui me portaient à le faire, mais ça arrive souvent que je les écris directement là. Des fois je les déplace par exemple. Pis les candidats je sais... quand je traite avec Envirodoq, les termes que je mets là sont toujours des termes qui sont dans le thésaurus. Je ne mets jamais d'autres termes. Non. Je pense que ça ne m'arrive jamais. Je ne mets jamais des termes qui ne sont pas dans le thésaurus, parce que je le connais. Pis quand j'en mets un, j'ai des forts doutes qu'il est là, je le vérifie mais, non, je mets toujours des termes du thésaurus. Si je sais qu'un terme, admettons qu'un document porte sur le biogaz. Je sais qu'il n'est pas dans le thésaurus, donc je vais tout de suite le mettre à la bonne place, dans les candidats.

E3 Là je me posais la question. *Biogaz*, je pensais qu'il l'était, mais je pense que dans notre nouvelle version il l'est. C'est peut-être ça. Ce terme là ça vaudrait la peine qu'il soit dans le thésaurus parce que dernièrement on a eu assez de documents là-dessus. Exemple, avec la ville de Montréal et la carrière Miron, on a eu un document en plusieurs volumes qui parlait justement de ça, *biogaz*.

L'évaluation. *Évaluation pour les emplacements... Évaluation des emplacements pour d'étangs aérés. Pour le traitement des eaux usées à Piedmont... Donc, site d'implantation... design des étangs.*

expé: Et tu retournes au document. Qu'est-ce que tu fais là?

E3 Je ne sais pas.

(consulte le thésaurus) Étangs, étangs, étangs aérés. Étangs pour le traitement des eaux usées. Étangs, eaux usées, d'aération, champs, traitement eaux usées, bactérie aérobie...

E3 Je suis allé voir les termes reliés je pense. À *étang aéré*, oui. Je suis allé voir les termes reliés pour voir si je ne pouvais pas avoir d'autres idées.

expé: Pour avoir d'autres idées?

E3 Pour être sûre d'avoir fait le tour de la question.

expé: Ça t'arrive souvent d'aller voir les termes alentour?

E3 Oui.

expé: Qu'est-ce que ça te permet de faire?

**E3** Ça me dit peut-être que ça, une personne pourrait chercher avec un terme qui est avoisinant, relié. Ça me fait poser la question, je devrais-tu le mettre ou pas. Des fois ça arrive qu'il y a des termes auxquels je n'aurais pas pensé, et je vais voir un descripteur et je vois tient on a cela dans le thésaurus! J'avais oublié ce terme là. Des fois je le rajoute dépendamment si le document en parle vraiment.

expé: As-tu tendance à prendre les mêmes descripteurs que t'utilises souvent ou si vos documents sont tellement variés que t'as une idée de l'ensemble..

**E3** Non , pour moi, je dois souvent utiliser les mêmes termes. Pas souvent dépendamment des documents. Mais si je suis partie sur la qualité de l'air plutôt que pollution air, ben je moi je suis partie sur qualité air, c'est ça que je vais employer.

Le volume, le design. Évaluation des sites.

expé: Tu retournes au document, qu'est-ce tu fais?

**E3** Je la fais là ma petite révision.

expé: Ce n'est pas de faire juste ta petite phrase.

**E3** Non, vu qu'il n'était pas épais, je l'ai refaite voir si j'avais rien oublié.

expé: est-ce que ça t'arrive des fois quand tu révises de tomber sur des nouveaux termes ou des trucs que t'avais pas vu?

**E3** Je ne sais pas. Ça peut arriver, mais je pense que c'est rare.

Contraintes environnementales, zonage... dézonage actuel... les zonages actuels sont identifiés... les différents sites, visibilité, contraintes d'évaluation, présences d'habitations...

**E3** Là j'essayais de bien faire mon travail, vu que je savais que j'étais observée.

expé: Mais as-tu l'impression que t'as fait comme si tu avais été en bas? Le même niveau...

**E3** Oui. Surtout ce document là, il n'était pas si compliqué que cela. Le résultat probablement qu'il aurait été la même chose, sauf que j'ai révisé un peu plus longtemps pour être sûre que ce soit bien fait. Parce que là, je pense que j'en rajoute plus.

expé: Est-ce tu révises de cette façon là normalement?

**E3** Je révise souvent. Mais des fois ça dépend du moment de la journée. Des fois quand il est rendu cinq heure moins le quart, je veux finir d'indexer ce document là et je ne veux pas m'en aller trop tard, je suis tannée. Des fois je me dis: ah! Ça va être beau comme cela. Mais je suis méticuleuse pour ça. Personnellement, je le suis, pas juste là-dedans. Je n'aime pas cela oublier des affaires ou ...

Contraintes physiques, contraintes techniques. Traitement retenu. Respect des exigences environnementales. Contraintes monétaires. Grille d'évaluation des sites... Évaluation détaillée des sites... Si... usine actuelle, le traitement par biofiltration a été proposé à l'étape 2 de l'étude préliminaire...

expé: Est-ce qu'il y a quelque chose là?

**E3** Je ne le sais pas. Non, je devais vérifier si j'avais pas oublier quelque chose.

expé: Comment tu fais pour t'apercevoir s'il te manque quelque chose?

**E3** Parce qu'on regarde vite quand même. On ne lit pas tout. Pis là en plus vu que je savais que j'étais observée, intérieurement j'étais un peu plus nerveuse. Donc j'avais un peu plus de misère à me concentrer que d'habitude. Là je ne voulais pas oublier quelque chose. C'est ça, vu qu'on regarde vite, on ne lit pas chaque phrase.

expé: La job ne vous permet pas de lire tout le document. As-tu l'impression que tu comprends quand même le document? Que t'as une bonne idée?

**E3** Ce document là, oui. Les autres aujourd'hui oui. Mais des fois ça arrive que quand je regarde, je ne sais pas de quoi il parle. Ça arrive des fois que pour certains documents, je vais fouiller dans les dictionnaires. Comme tantôt, il y avait un terme. Si j'avais été dans mon bureau, j'aurais regardé (comme pour *exhaure de l'eau*) Je vais dans les dictionnaires, les encyclopédies pour être sûre. Quand c'est des documents anglais aussi, je suis moins habile en anglais. C'est plus rare dans Envirodoq c'est assez rare. Mais la collection générale, c'est quasiment, la plupart sont en anglais..

expé: Vous les indexez avec des termes en français?

**E3** Oui. L'autre indexeur c'est ce qu'il fait. Mais vu que se sont des monographies, il y a souvent les vedettes-matières. Il vérifie beaucoup avec l'index des vedettes-matières de LAVAL, il y a un document avec les termes anglais et français. Il vérifie et après, il sait que le terme anglais est traduit comme cela dans notre thésaurus.

expé: Une question plus générale. Qu'est-ce qui se passe quand tu indexes comparativement à une lecture ordinaire?

**E3** Ça dépend du genre de document. Si je lis un roman, je ne pense à rien, j'embarque dans l'histoire. Si je lis des affaires sur des plantes, ou... je suis portée à aller voir ce qui m'intéresse. Je regarde vite quand même. Ma manière de lire vite peut-être qu'elle est la même, j'essais de visualiser dans ma tête une page pis de voir de quoi ça parle. Et si je vois quelque chose qui m'intéresse, je vais m'arrêter et je vais lire plus attentivement. Mais là que je fais l'indexation, ma méthode de lire est peut-être un peu pareille, je vais vite. Et dans ma tête j'ai comme mon thésaurus en tête. J'essais de ne pas délaissier des parties au détriment de d'autres. Mais si ça n'a pas de lien, si les parties sont pas refléter dans le thésaurus, peut-être qu'inconsciemment je vais être porté plus à les délaissier à moins que ça soit vraiment important. Mais ça peut arriver peut-être.

expé: C'est quoi vraiment important?

**E3** Comme le *biogaz*. Même s'il n'est pas dans le thésaurus, je peux pas le laisser de côté. Je ne peux vraiment pas le laisser tomber. Je ne sais pas, peut-être que si *contraintes* avait été dans le thésaurus, je l'aurais peut-être mis. Dans ma tête j'aurais pensé, tiens contraintes on a ça je vais mettre ça. Mais vu qu'il ne l'est pas...

C'est des *contraintes*, mais dans le fond ça rapport à *l'évaluation des sites*. Oui, là j'ai mis *contraintes*. Mais *contraintes* c'est aussi vague que *l'évaluation*.

expé: ce sont des beaux mots mais, tout seul...

**E3** Il n'y a personne qui va chercher là dessus.

<p>a été suggérée à l'étape 2 se situe... Emplacement de l'usine, bâtiment, garage...  débosselage, recyclage... partie... tamisé, zone inondable. Variante, biofiltration. Trois.  Trois, c'est évaluation des sites. Contraintes d'évaluation. Contraintes monétaires...</p>
--

expé: Tu repasses les contraintes.

E3 Là je sais pas si je suis allé voir la table des matières?

expé: T'es rendue au point trois. Tu les nommes (*les contraintes*).

Table des matières. Les contraintes et l'évaluation. Contraintes. Évaluation des sites. Contraintes... contraintes... contraintes...

E3 Là j'ai dû penser contraintes est-ce que je le mets ou non. Qu'est-ce que je fais avec contraintes?

expé: En dernier au niveau de ta vérification comme ça, tu traites ce concept là. Était-ce un aspect que tu croyais que t'aurais pu avoir oublier? Je ne sais pas, moi j'essaie de te pose des questions. Je peux être complètement dans les patates. Pour que tu le sortes...

E3 Je ne le sais pas. C'est inconscient.

E3 On est influencé par le vocabulaire du livre. Et je suis sûre que la personne à la référence est influencée par la formulation des questions. Elle cherche de telle manière parce que la personne lui a demandé de telle manière. C'est pour ça que c'est dur de se rejoindre.

Non. Je vais arrêter là. Ça fait que ça va être les descripteurs principaux ça va être: **évaluation, l'emplacement... d'étang aéré, traitement eaux usées** en descripteur secondaire, étant donné que c'est plus général. Traitement eaux usées. Pis en géo, **Piedmont**.

Annexe D

Sections consultées du document

***Étang aéré***

**SOCIÉTÉ QUÉBÉCOISE D'ASSAINISSEMENT DES EAUX**

**P.A.E.Q.**

**ST-SAUVEUR-DES-MONTS**

**ST-SAUVEUR PAROISSE**

**PIEDMONT**

**ÉTUDE DE SITES D'IMPLANTATION**

**D'ÉTANGS AÉRÉS**

**CONSULTANTS**

**CEDEGER LTÉE**  
Interception

**GENDRON LEFEBVRE INC.**  
Traitement et réhabilitation

Octobre 1990

SOCIÉTÉ QUÉBÉCOISE D'ASSAINISSEMENT DES EAUX

P.A.E.Q.

ST-SAUVEUR-DES-MONTS

ST-SAUVEUR PAROISSE

PIEDMONT

ÉTUDE DE SITES D'IMPLANTATION D'ÉTANGS AÉRÉS

À PIEDMONT

Alain Gadbois, ing. jr. M.Sc.A.  
Ingénieur  
Génie Urbain

Chantal Morasse, ing. M.Sc.A.  
Chef de Service  
Génie Urbain

TABLE DES MATIÈRES

	<u>Page</u>
1. INTRODUCTION	1
2. ÉTANGS AÉRÉS, SITES D'IMPLANTATION ET CRITÈRES DE DESIGN	1
2.1 Sites d'implantation	1
2.2 Design des étangs	2
3. ÉVALUATION DES SITES	5
3.1 Contraintes d'évaluation des sites	5
3.1.1 Contraintes environnementales	5
3.1.2 Contraintes physiques	6
3.1.3 Contraintes techniques	6
3.1.4 Contraintes monétaires	8
3.2 Grille d'évaluation des sites	9
3.3 Évaluation détaillée des sites	9
3.3.1 Site usine actuelle	9
3.3.2 Variante - étangs aérés facultatifs	12
3.3.3 Site Lutfy	13
3.3.4 Site Méandre	16
3.3.5 Boysfarm	20
3.3.6 Ferme Monfette	23
4. CONCLUSION	25

## 1. INTRODUCTION

Les municipalités de Piedmont, Saint-Sauveur-des-Monts et St-Sauveur Paroisse se sont regroupées dans le but d'effectuer un traitement commun de leurs eaux usées se déversant dans la Rivière du Nord. Une épuration par biofiltration ou par étangs facultatifs aérés leur a été proposée à l'étape 2 de l'étude préliminaire. L'emplacement déterminé pour les étangs aérés étant situé dans une zone à vocation résidentielle, la Société Québécoise d'Assainissement des Eaux a mandaté le consortium Gendron Lefebvre - CEDEGER afin d'identifier des sites alternatifs d'implantation de l'usine d'épuration. Ces sites seront déterminés parmi 5 différents emplacements potentiels situés dans la municipalité de Piedmont. Chaque site sera jugé dans le présent document en fonction de contraintes environnementales, physiques, techniques et monétaires. Le présent rapport résume les avantages et les inconvénients de chacun. Il fait également état des valeurs monétaires des terrains ainsi que des dommages potentiels à défrayer pour l'acquisition des sites. Aucune évaluation des coûts de construction n'y est présentée.

## 2. ÉTANGS AÉRÉS, SITES D'IMPLANTATION ET CRITÈRES DE DESIGN

### 2.1 Sites d'implantation

Les cinq sites d'implantation à Piedmont sont connus sous les appellations suivantes:

- 1) Usine actuelle.
- 2) Lutfy.
- 3) Méandre.
- 4) Boysfarm.
- 5) Ferme Monfette.

#### 4. CONCLUSION

Cette étude a permis d'analyser cinq différents sites pour l'implantation d'étangs aérés et d'une usine de biofiltration dans la municipalité de Piedmont.

L'implantation d'étangs aérés nécessitant environ 10 hectares en plus de 25 hectares comme zone de protection, la grille d'évaluation fait voir qu'aucun des cinq sites ne permet l'utilisation d'une telle superficie sans soulever certaines contraintes environnementales, physiques, techniques ou monétaires.

L'implantation d'étangs sur le site de l'usine actuelle nécessite l'expropriation de 12 maisons. De plus, le territoire est nettement situé en zone inondable. Cette variante est donc à rejeter.

L'implantation d'une usine de biofiltration sur le site adjacent à l'usine actuelle, telle que proposé à l'étape 2 de l'étude préliminaire, n'exige que peu de superficie. Un dézonage est cependant requis ainsi qu'une dérogation de zone inondable. Ces deux contraintes peuvent toutefois être évitées en établissant la nouvelle usine sur le site même de l'usine actuelle.

Le site Boysfarm présente une topographie accidentée compliquant grandement la construction et l'opération des étangs. La superficie restreinte, la présence de ruisseaux, l'accès difficile, sont également des contraintes non négligeables. L'utilisation de ce site compromettrait grandement l'installation d'un terrain de golf.

Le site Lutfy ne présente aucune contrainte technique ni physique. La grande visibilité et le zonage résidentiel demeurent les seules contraintes.

Le site Méandre ne permet pas l'implantation d'étangs aérés ayant un temps de contact de 15 d. Une solution mixte incorporant des étangs aérés, une filtration sur sable et une désinfection U.V. doit être envisagée. L'utilisation de ce site se complique également par l'obligation de procéder à des expropriations inter-municipales ainsi que par la construction en partie en zone inondable, qui demeure à être confirmée par arpentage.

Le site Ferme Monfette ne présente à prime abord aucune contrainte technique majeure. La topographie et la superficie seraient acceptables mais demeurent à confirmer par arpentage. La visibilité du site est moins importante que celle du site Lutfy mais l'impact négatif sur le développement résidentiel présent et futur aux alentours serait important. L'utilisation de ce site se complique également par l'obligation de procéder à un dézouage.

En conclusion, la grille d'évaluation fait voir qu'aucun site ne ressort du groupe de façon évidente pour l'implantation d'étangs aérés. Tous les sites présentent en effet diverses contraintes à tenir en ligne de compte pour le choix final du site et du type de traitement.

Annexe E

Sections consultées du document

***Biogaz***

7271-2400



VILLE DE MONTRÉAL  
Services des travaux publics  
Module Génie de l'environnement

---

Centre de tri et d'élimination des déchets.  
Élimination du biogaz.  
Sélection d'une technologie et évaluation des impacts.

Rapport préliminaire

Janvier 1992



**TECSULT**

**Tecsult Inc.**

experts-conseils

85, RUE STE-CATHERINE OUEST, MONTRÉAL (QUÉBEC) CANADA

## SOMMAIRE

Tecsult Inc. a été mandaté par la Ville de Montréal pour sélectionner une technologie adéquate de destruction du biogaz produit par le site d'enfouissement du Centre de tri et d'élimination des déchets (CTED). Ce mandat prévoit également une évaluation des impacts causés par cette technologie.

Ce rapport présente la démarche suivie pour atteindre ces objectifs. La première étape consiste à faire la synthèse des caractérisations effectuées à ce jour sur la composition du biogaz. Plusieurs scénarios de composition ont été retenus pour le dimensionnement des équipements. Toutefois, les études ont permis de conclure que la composition typique qui doit être utilisée pour la conception est la suivante:

Méthane (CH <sub>4</sub> )	40%
Gaz carbonique (CO <sub>2</sub> )	38%
Oxygène (O <sub>2</sub> )	6%
Azote (N <sub>2</sub> )	16%

La concentration de composés en traces dans le biogaz a également été estimée à partir d'une synthèse des études effectuées à ce jour pour le compte de la Ville. Une revue de littérature a permis de compléter les informations manquantes. Près de 35 substances ont ainsi été retenues pour l'évaluation des impacts sur la qualité de l'air. Les substances étaient retenues soit en fonction de leur importance quantitative, soit en fonction des exigences du Règlement 90 de la CUM. Les composés organiques volatils (COV) considérés prioritaires en Californie par le Calderon Bill et ceux observés dans une étude portant sur plusieurs sites d'enfouissement américains ont également été retenus.

La production de biogaz ira en augmentant pour atteindre un maximum d'environ 380 millions de m<sup>3</sup>/an de biogaz, selon les estimations de Tecsult pour l'année 1993. Les équipements devront pouvoir éliminer tout le biogaz capté au CTED avant et après la période de démarrage de l'usine de valorisation énergétique ainsi que le surplus que cette dernière ne pourra convertir en énergie. La production de biogaz diminuera, par la suite, de façon exponentielle pendant plusieurs dizaines d'années.

Outre les contraintes énergétiques liées à la concentration de méthane dans le biogaz et à la quantité de biogaz à éliminer, d'autres critères de conception doivent être considérés. La présence de composés organiques volatils (COV) dans le biogaz entraîne la nécessité de prendre des mesures pour assurer leur destruction. À ce propos, l'expérience démontre qu'un temps de rétention des gaz de combustion de 1 seconde à une température de 1000° C est suffisante pour assurer une destruction des COV. L'air doit être présent en excès pour éviter la formation de monoxyde de carbone. Le mélange entre l'air et les gaz de combustion doit être suffisant pour éviter la formation de canaux d'écoulement préférentiels réduisant l'efficacité de combustion. Finalement, la technologie retenue doit, tout en respectant ces critères, s'avérer fiable et économique autant au moment de l'exploitation que de la construction.

Le processus de sélection a permis d'identifier trois technologies disponibles pour éliminer le biogaz, soit les torchères à flamme visible, les torchères à flamme invisible et les incinérateurs. Les torchères à flamme visible s'avèrent les moins coûteuses et les plus simples d'opération. Toutefois, elles ne peuvent rencontrer les critères environnementaux de temps de rétention et de température. Les incinérateurs remplissent les exigences environnementales mais s'avèrent fort coûteux. De plus, très peu d'installations de ce type sont actuellement en exploitation pour l'élimination du biogaz.

Les torchères à flamme invisible permettent de rencontrer tous les critères tout en offrant un coût largement inférieur à celui des incinérateurs. Cette technologie est d'ailleurs la plus répandue en Californie. La Ville de Toronto en a d'ailleurs installé dans deux de ses centres d'enfouissement.

Une demande d'information auprès de deux fournisseurs différents a permis de vérifier la disponibilité de torchères de ce type. Les dimensions typiques pour les besoins du CTED sont de 15 m (50 pi) de haut par 3.7 m (12 pi) de diamètre. Un bilan de masse et d'énergie a permis de constater que, pour le débit maximal de biogaz, huit torchères seraient requises et que le temps de rétention de 1 seconde à 1000° C pourrait facilement être respecté.

Les impacts de cette technologie ont été évalués en fonction des préoccupations du public telles qu'identifiées par une revue de presse et par consultation auprès des intervenants. Ces préoccupations sont:

- l'impact sur la qualité de l'air;
- le bruit du système de compresseurs;
- les risques de déflagration;
- les odeurs;
- le transport aérien.

L'évaluation des impacts est initiée par une description du milieu environnant le site d'enfouissement. Ce milieu s'avère urbain, fortement résidentiel et commercial. Des parcs, des écoles, des hôpitaux et quelques jardins communautaires y sont observés.

L'impact sur la qualité de l'air est évalué de deux façons. La première consiste à comparer les émissions aux exigences du Règlement 90 de la CUM sur la qualité de l'air. Pour toutes les substances identifiées, les concentrations dans les gaz de combustion sont bien en deça des normes.

La seconde évaluation de l'impact sur la qualité de l'air est effectuée en modélisant la dispersion des polluants dans l'atmosphère à l'aide du logiciel ISCST développé par l'USEPA. Les concentrations dans l'air obtenues sont comparées aux niveaux de fond actuels et comparés aux normes du Règlement 90. L'apport maximal par rapport à la norme est de l'ordre de 9% pour le SO<sub>2</sub>, les NO<sub>x</sub> et les particules. Le CO n'entraîne dans aucun cas un apport supérieur à 1%.

Le bruit produit par les compresseurs sera minimisé par des enceintes acoustiques et par des silencieux. À l'intérieur du bâtiment, le bruit respectera les normes du Règlement sur la Qualité du milieu de travail, soit 90 dBA. À l'extérieur du bâtiment, le bruit n'excédera pas 45 dBA, respectant ainsi la norme de la Ville de Montréal.

Les risques de déflagration seront minimisés par l'installation de systèmes de détection du méthane, par des équipements de protection-incendie et de ventilation adéquats. Le méthane est inflammable à des concentrations entre 5% et 15% dans l'air. Différents systèmes, décrits dans le rapport, seront installés pour prévenir cette plage de concentrations.

Le contrôle des odeurs est assuré par une combustion efficace. La combustion est reconnue comme la meilleure méthode de destruction des odeurs. À plus de 1000° C, les produits malodorants sont tous détruits. Les émissions fugitives de biogaz, si elles se produisent, seront insuffisantes pour entraîner des odeurs, à moins d'un bris complet d'une conduite.

Les professionnels de Tecsuit ont rencontré les responsables de la navigation aérienne de Transport Canada pour la région de Québec. Ceux-ci ont confirmé que l'emplacement prévu des torchères se trouve dans l'axe d'un corridor aérien. Toutefois, ce corridor se situe à haute altitude pour la plupart des avions utilisant la piste concernée à l'aéroport Dorval et l'impact devrait être marginal.

De petits avions circulent fréquemment au-dessus de ce secteur. Transport Canada désire vérifier l'impact des émissions de la torchère sur ce type de navigation. Les résultats de cette vérification seront inclus dans le rapport final de Tecsuit.

## TABLE DES MATIÈRES

		<u>Page</u>
1.	INTRODUCTION .....	1
	1.1 Contexte .....	1
	1.2 Mandat .....	2
	1.3 Travaux réalisés .....	2
2.	CARACTÉRISATION DU BIOGAZ .....	4
	2.1 Objectif .....	4
	2.2 Historique des campagnes d'échantillonnage .....	5
	2.2.1 Campagne de mai 1989 - SCS Engineers .....	5
	2.2.2 Campagne de mai 1989 - Université Laval .....	6
	2.2.3 Campagne de janvier 1989 - ÉCO-recherches .....	9
	2.2.4 Campagne de mai 1988 - Novalab .....	9
	2.3 Élaboration d'une caractérisation globale .....	11
	2.3.1 Critères de validation des résultats de laboratoire .....	11
	2.3.2 Cadre juridique .....	12
	2.3.3 Revue de littérature .....	13
	2.4 Synthèse des campagnes d'échantillonnage et sélection des substances les plus importantes .....	16
	2.4.1 Les composés majeurs .....	16
	2.4.2 Les composés organiques volatiles (COV) .....	18
3.	ÉVALUATION DES DÉBITS DE BIOGAZ .....	20
	3.1 Génération de biogaz .....	20
	3.2 Variation dans le temps de la composition du biogaz .....	21
4.	SÉLECTION D'UNE TECHNOLOGIE D'ÉLIMINATION .....	22
5.	CRITERES DE CONCEPTION DES TORCHÈRES À FLAMME INVISIBLE ..	24
	5.1 Paramètres de conception .....	24

## TABLE DES MATIÈRES

(suite)

	<u>Page</u>
5.1.1	Température minimale d'opération et temps de séjour des gaz de combustion ..... 24
5.1.2	Excès d'air ..... 25
5.1.3	Débit de conception ..... 26
5.2	Bilan d'énergie ..... 26
5.2.1	Bilan d'énergie sur une torchère type ..... 28
5.2.2	Plage de sensibilité sur les compositions ..... 28
5.2.3	Besoin éventuel en combustible d'appoint ..... 31
5.3	Bilan de masse ..... 31
5.3.1	Bilan de masse sur une torchère type ..... 31
5.3.2	Effet de la composition du biogaz sur la vitesse de sortie des gaz de combustion et sur le temps de rétention ..... 33
5.4	Synthèse ..... 33
6.	ÉVALUATION DES IMPACTS ..... 36
6.1	Préoccupation du public ..... 36
6.2	Qualité de l'air ..... 36
6.2.1	Mécanismes de la dispersion des polluants ..... 37
6.2.2	Modèle utilisé ..... 40
6.2.2.1	Choix du modèle ..... 40
6.2.2.2	Description du modèle ISCST ..... 42
6.2.3	Paramètres de modélisation ..... 46
6.2.3.1	Description de l'aire d'étude ..... 46
6.2.3.2	Les données techniques ..... 48
6.2.4	Résultats de la modélisation ..... 54
6.2.4.1	Résultats sur l'ensemble de la grille ..... 55
6.2.4.2	Résultats aux postes de la CUM ..... 56
6.2.4.3	Résultats aux récepteurs sensibles ..... 61

**TABLE DES MATIÈRES**  
(suite)

	<u>Page</u>
6.2.5 Modèle du règlement 90 de la CUM .....	61
6.3 Contrôle du bruit du bâtiment des compresseurs .....	62
6.3.1 Étude .....	64
6.3.2 Analyse .....	64
6.4 Risques de déflagration .....	65
6.5 Odeurs .....	66
6.6 Transport aérien .....	67
6.7 Autres impacts .....	68
6.7.1 Main-d'oeuvre .....	68
6.7.2 Camionnage et circulation automobile .....	68
6.7.3 Contrôle des eaux usées .....	68
6.7.4 Coût du projet de pompage et brûlage du biogaz .....	69
7. CONCLUSION .....	70

## 1. INTRODUCTION

### 1.1 Contexte

L'enfouissement des déchets au Centre de Tri et d'Élimination des Déchets (CTED) a débuté en 1968 par la compagnie Miron qui exploitait alors la carrière. La ville de Montréal a pris possession de l'ancienne carrière en 1988 et supervise depuis lors les activités d'enfouissement. Actuellement, le site contient près de 30 millions de tonnes de déchets et pourra en contenir près de 35 millions à sa fermeture, prévue en 1994.

Le biogaz provient de la décomposition anaérobie des déchets enfouis, ces déchets peuvent générer jusqu'à 43 000 mètres cubes par heure de biogaz (Castiello et Dubois, 1991). Le biogaz est surtout composé de méthane et de bioxyde de carbone. Il contient également en trace plusieurs composés organiques volatils. La qualité du biogaz ainsi que son taux de production peut varier selon plusieurs facteurs: l'âge des déchets, l'état du réseau de captage, l'état des puits et du recouvrement final, l'humidité et la température des déchets, la pression atmosphérique, le mode d'opération du site, etc.

Le biogaz constitue une nuisance environnementale. Le méthane qu'il contient est inflammable et reconnu pour favoriser l'effet de serre. Des odeurs nauséabondes résultent de sa présence. Le biogaz contient des traces de composés organiques volatils qui peuvent s'avérer toxiques ou cancérigènes.

La migration du biogaz est donc contrée par un réseau de captage constitué de puits et de conduites souterraines. Par le biais de compresseurs, le biogaz est acheminé vers un système d'élimination. Une partie du biogaz servira à alimenter une usine de cogénération d'électricité. Le surplus devra être brûlé dans un système adéquat. L'objet

de ce rapport est donc de présenter les critères de conception de ce système et les impacts qui en résultent.

## 1.2 Mandat

Le mandat confié à Tecsalt Inc. par la Ville de Montréal comporte les éléments suivants:

1. Sélectionner une technologie adéquate pour détruire la totalité du biogaz capté au CTED en respectant les priorités et contraintes imposées par la mise en route du système de captage.
2. Évaluer les émissions et les retombées atmosphériques de cette technologie.
3. Comparer les émissions et les retombées atmosphériques prévues avec le Cadre juridique en vigueur ainsi qu'avec les plus récentes exigences canadiennes et californiennes.

## 1.3 Travaux réalisés

Quelques points se sont ajoutés au mandat initial, soit une étude sur le niveau sonore des compresseurs, une étude sur les risques de déflagration et une étude sur les odeurs. Puisque les installations du CTED se trouvent à proximité de l'aéroport Dorval, les professionnels de Tecsalt ont rencontré les responsables de Transport Canada pour discuter de l'effet des émissions de gaz chauds au-dessus de corridors aériens menant à l'aéroport Dorval.

Annexe F

Sections consultées du document

***Compostage***



RESIDUAGES  
COMPOSTAGES



Agence de Bassin  
SEINE-NORMANDIE



Agence de Bassin  
LOIRE-BRETAGNE

Décembre 1984

*Roger Bessier*

MINISTÈRE DE L'ENVIRONNEMENT

# COMPOSTAGE DES BOUES RÉSIDUAIRES URBAINES

---

Agence de bassin Seine-Normandie  
Agence de bassin Loire-Bretagne

Décembre 1984

## PRÉFACE

*Il est particulièrement agréable, en tant que directeurs d'agences de bassin, de présenter un document qui est le fruit d'une collaboration entre des personnes d'horizons divers mais qui ont toutes un même objectif : celui de mettre leurs connaissances et leurs compétences au service des collectivités, pour les aider à résoudre les problèmes auxquels elles sont confrontées en matière de protection de l'environnement.*

*L'importance de plus en plus grande accordée à la valorisation des résidus de l'épuration, les boues, a conduit en effet les agences de bassin à favoriser une réflexion sur les techniques qui pouvaient être mises à la disposition des élus. C'est ainsi que diverses études, expériences et publications ont été consacrées aux différentes solutions envisageables : la valorisation en agriculture des boues à l'état liquide ou après déshydratation, valorisation énergétique des boues par utilisation du gaz méthane, issu de la digestion des boues, etc... Restait le compostage, procédé ancien bien connu, utilisé dans différents domaines dont le traitement des ordures ménagères, mais qui n'avait reçu en France que très peu d'applications dans le traitement des boues d'épuration.*

*Les résultats des réalisations à l'étranger et des premières expériences françaises sont suffisamment satisfaisants pour qu'il paraisse nécessaire de faire le point sur les techniques actuellement en présence sur le marché. Il ne s'agit pas ici de recherche fondamentale, mission dévolue à d'autres organismes, mais simplement de dresser un « état de l'art » afin d'apporter aux décideurs et aux maîtres d'œuvre les informations techniques dont ils ont besoin.*

*Il est toujours bon de remettre ces techniques contemporaines dans le contexte qui a permis leur développement. En effet, sans remonter à l'origine du compostage, qui est certainement une très ancienne pratique culturelle, on sait que c'est Sir A. Howard, qui dès 1920 a posé les premières bases d'un compostage permettant de résoudre les problèmes d'hygiène posés par les déjections humaines et animales. Dès cette époque, des brevets ont été pris. Durant les années 50 et 60, une modification des comportements entraîne le rapide foisonnement des ordures ménagères. Des procédés apparaissent qui apportent une solution à ce problème, et déjà se manifeste un intérêt pour le compostage des ordures avec les boues de stations d'épuration. Mais à cet enthousiasme, succède un certain désenchantement en raison de la faible valeur des composts obtenus dans une période où le coût des engrais était relativement bas.*

*Dans les années 70, l'augmentation du coût de l'énergie (qui rend plus coûteuses certaines techniques d'élimination des boues, comme l'incinération), les préoccupations environnementales (à l'égard des décharges en particulier), le coût des engrais (lié au coût de l'énergie) et le regain d'intérêt du monde agricole pour un apport en matière organique, à des sols de plus en plus pauvres en humus, entraînent un essor des solutions valorisantes aux problèmes des déchets. C'est aux États-Unis, en Allemagne et en Suisse que le compostage des boues entre d'abord en application. En Suisse en particulier, les stations d'épuration sont installées à proximité des usines de traitement d'ordures ménagères de façon à permettre un compostage commun des 2 déchets. Puis, en Allemagne et en France, se développent les techniques utilisant sciures et écorces.*

*Au milieu de la décennie 80, il convient de présenter un premier bilan, technique et économique. C'est l'objet de ce document.*

*Nous tenons ici à remercier les sociétés qui ont aidé à la rédaction de ce document, à savoir :*

*AGRO-DÉVELOPPEMENT, GIE ANJOU RECHERCHE, O.T.V., TERNOIS ÉPURATION, TRIGA, SAUR et spécialement MM. MOUSTY et MARTEL qui dans le cadre de leur thèse de doctorat, ont apporté un concours très précieux aux techniciens des agences de bassin Loire-Bretagne et Seine-Normandie.*

*Enfin, il convient de souligner que la réalisation de ce document n'aurait pu voir le jour sans l'étroite collaboration établie entre l'ANRED et les agences de bassin.*

J.C. VINÇONNEAU

Directeur de l'agence de bassin Loire-Bretagne

C. FABRET

Directeur de l'agence de bassin Seine-Normandie

## SOMMAIRE

	<i>Pages</i>
Terminologie et abréviations	6
Liste des tableaux	7
Liste des photos	8
Liste des figures	9
Avertissement	10
Introduction	11
1. — Pourquoi le compostage ?	13
1.1. — Les boues et les facteurs conditionnant leur valorisation agricole	
1.1.1. — Incidence du traitement sur la composition des boues	
1.1.2. — Contraintes liées aux boues	
1.2. — Intérêt du compostage	
2. — Le compostage : principes et technologies adaptées au traitement des boues	
2.1. — Définition	
2.2. — Fondements du compostage	
2.2.1. — Principe	
2.2.2. — Conditions requises pour le compostage des boues	
2.3. — Le processus et son suivi	
2.3.1. — Les températures	
2.3.2. — Les autres paramètres	
2.4. — La maturation	
2.5. — L'évaluation de l'efficacité du compostage	
2.5.1. — La destruction des germes pathogènes	
2.5.2. — L'élimination d'eau consécutive à l'aération et aux phénomènes de thermogénèse	
2.5.3. — La valeur agronomique du produit obtenu	
3. — Les techniques développées en France	29
3.1. — Classification	
3.1.1. — Par la technologie employée	
3.1.2. — En fonction du moyen d'aération	
3.1.3. — En fonction des applications actuellement développées	
3.2. — Le compostage « accéléré » en bioréacteurs	
3.2.1. — Le procédé BAV	
3.2.2. — Le procédé TRIGA	
3.2.3. — Commentaire général	
3.3. — Le compostage en tas	
3.3.1. — Essais du C.R.E.A.T.E.	
3.3.2. — Le procédé BELTSVILLE - RUTGERS	
3.3.3. — Le procédé BIOTERRA	
3.4. — Le compostage en andains	
3.4.1. — Le C400	
3.4.2. — Le Scarabée	
3.4.3. — Remarque générale sur le compostage en andains	
3.4.4. — Les techniques rustiques : l'épandeur à fumier	
3.5. — Compostage avec retournement et aération forcée	

4. - Le projet de compostage : l'aspect économique	39
4.1. - L'aspect commercial	
4.1.1. - L'agent structurant	
4.1.2. - Les débouchés pour le compost	
4.2. - L'établissement du projet technique	
4.2.1. - Renseignements à fournir aux constructeurs	
4.2.2. - Les renseignements à demander	
4.2.3. - L'estimation des coûts	
4.3. - Quelques éléments de comparaison	
5. - Compostage de boues liquides - Application aux stations d'épuration de 2 000 à 15 000 équivalents-habitants de capacité - Filières n° 1 et 2	49
5.1. - Description des essais	
5.2. - Préparation des boues	
5.3. - Préparation de la paille	
5.4. - Compostage	
5.5. - Ecoulement, commercialisation et bilan économique	
5.6. - Autres données sur la filière : expérience de SEES - 61	
6. - Filières de compostage des boues déshydratées - Applications de 25 000 à 500 000 équivalents-habitants	59
6.1. - Le compostage « accéléré » en bioréacteurs	
6.1.1. - Filière n° 3 : Procédé TRIGA/Hygiénisateur du SIVOM de la presqu'île d'ARVERT	
6.1.2. - Filière n° 4 : BAV - NANTES Sud	
6.2. - Compostage avec retournement d'andains	
6.2.1. - Filière n° 5 : Compostage avec retournement d'andains, BLOIS (41), 70 000 équivalents-habitants	
6.2.2. - Filière n° 6 : SCARABEE/SABLE	
6.3. - Compostage en tas avec ventilation forcée	
6.3.1. - Filière n° 7 : BIOTERRA/SURGÈRES (17)	
6.3.2. - Filière n° 8 : BELTSVILLE - RUTGERS/OZOIR-LA-FERRIÈRE	
6.4. - Compostage associant retournement et ventilation forcée	
6.4.1. - Filière n° 9 : Procédé SILODA	
6.4.2. - Remarque sur le procédé DANO	
6.5. - Lombricompostage	
7. - Conclusion	99
8. - Annexes	101
8.1. - Le compostage de boues liquides - Application aux unités de 2 000 à 15 000 équivalents-habitants de capacité	
8.1.1. - Description des essais	
8.1.2. - Réception et stockage des produits à mélanger (S1)	
8.1.3. - Conditionnement des boues/Déshydratation par filtration (S2)	
8.1.4. - Fonctionnement du support carboné (S3)	
8.1.5. - Mélange (S4)	
8.1.6. - Compostage (S5)	
8.1.7. - Maturation (S6)	
8.1.8. - Post broyage (éventuel) (S7)	
8.1.9. - Ecoulement et commercialisation (S8) - bilan économique	
8.2. - Test de mesure de la maturité	120
8.2.1. - Test : azote soluble	
8.2.2. - Test des sulfates	
8.2.3. - Test de croissance du cresson (variété ALENOIS)	
8.3. - Bibliographie	122

## INTRODUCTION

Depuis 25 ans, un effort important a été accompli dans le domaine de l'assainissement des collectivités locales. Cet équipement collectif (réseau plus station d'épuration) est devenu un symbole de la protection de l'environnement. Mais peu d'usagers savent que la station d'épuration ne fait que concentrer la pollution, qui passe de l'eau — épurée — vers une solution plus concentrée, les boues.

Or, ces boues s'avèrent, elles aussi, difficiles à éliminer. Le traitement des boues représente, en effet, une charge importante pour la collectivité : au moins 30 % de l'investissement de la station d'épuration, un pourcentage du même ordre des dépenses de fonctionnement, pour obtenir un produit qui va encombrer les décharges ou dont l'incinération sera prohibitive. Valoriser ce produit répond donc à un besoin. Cela permet de diminuer au maximum le coût d'exploitation de la station tout en contribuant au recyclage des produits fertilisants contenus dans la boue.

A l'autre bout de la chaîne, ce raisonnement est par ailleurs conforté par la situation de certains sols sur le plan agronomique. A l'apport d'engrais, vient s'ajouter le besoin en humus pour des raisons d'amélioration ou de maintien de texture du sol. On a donc vu peu à peu se multiplier les opérations de valorisation des boues en agriculture sous diverses formes : épandage liquide, épandage à l'état solide ou pâteux. Ces solutions ont leurs avantages et leurs inconvénients, qui ont fait l'objet de nombreuses publications depuis 10 ans bientôt.

Le compostage s'est progressivement dégagé comme étant une solution qui permette de pallier certains inconvénients tout en gardant l'avantage principal : un apport de fertilisants et d'humus. Au début, et contrairement à ce qui s'est passé aux États-Unis par exemple, ce sont de grandes installations de compostage (pour 50 000 Eq. Hab. (1)) qui sont apparues en France. Depuis, en une dizaine d'années, de nouvelles techniques permettent d'envisager le compostage pour des unités plus petites (50 000, puis 25 000 Eq. Hab.). Enfin, à la suite des travaux du C.R.E.A.T.E., des résultats ont montré que sous certaines conditions, le compostage était concevable pour des stations de 2 000 à 5 000 Eq. Hab.

Le compostage apparaît donc comme une solution fort intéressante pour la valorisation agricole des boues, solution qui est encore sous-utilisée en France. Mais ce n'est pas une solution miracle, en particulier au niveau des petites stations rurales pour lesquelles la valorisation en l'état reste la solution très souvent la mieux adaptée.

L'objet du présent document est de décrire succinctement cette panoplie de techniques à la disposition des décideurs, afin de faciliter leurs choix. Il n'a pas la prétention de faire un point exhaustif sur le sujet, mais de tirer les premières conclusions des résultats déjà obtenus. Sa composition est la suivante :

- le premier paragraphe donne les raisons du compostage ;
- le deuxième est une brève définition du compostage ;
- le troisième présente les diverses techniques en présence ;
- le quatrième paragraphe s'intéresse à l'aspect économique qui semble fondamental pour l'avenir de cette technique ;
- les cinquième et sixième sont plus technologiques et donnent les principales caractéristiques des filières en présence ;
- des annexes permettent au lecteur de trouver les descriptifs plus détaillés, des méthodes de mesure utilisées couramment pour la qualité du compost, la liste des principaux ouvrages et publications utilisés pour la rédaction du document. Le lecteur y trouvera également un rapport qui peut se lire indépendamment sur le compostage des boues liquides.

(1) Eq. Hab. = Equivalent Habitant.

Annexe G

Sections consultées du document

***Énergie***

2-1162803

UNIVERSITE DU QUEBEC

RAPPORT DE RECHERCHE

PRESENTE A

L'UNIVERSITE DU QUEBEC A MONTREAL

COMME EXIGENCE PARTIELLE

DE LA MAITRISE EN SCIENCES DE L'ENVIRONNEMENT

PAR

Bruno Desbois  
géographe

ETUDE DE LA POSSIBILITE D'UTILISER LES ENERGIES  
RENOUVELABLES, (EOLIENNE, SOLAIRE ET DE LA BIOMASSE)  
AU VILLAGE DE NDIA GOUREYE AU SENEGAL

FEVRIER 1985

RESUME

Le village de Ndia Gouréyé est situé au nord du Sénégal, dans la zone climatique sahélienne. Le climat est caractérisé par une très forte variation des pluies et par des précipitations annuelles comprises entre 400 et 600 mm de hauteur d'eau. La période de l'hivernage y dure environ 3 mois. Pendant la saison sèche et notamment aux mois de décembre et janvier, un vent froid et sec, le harmatan, y souffle du N-E avec beaucoup de vigueur. Le sol est surtout constitué d'éléments sablonneux: les sols dior, sur lesquels poussent facilement l'arachide. L'ensoleillement du site y est très important, ce qui accroît le phénomène de désertification, très important dans cette région sahélienne. La végétation naturelle de la région est constituée d'un réseau clairsemé d'acacia, de soump, de baobab et de tamarinier qui, progressivement depuis de nombreuses années, a laissé place aux champs d'arachides, de mil, de sorgho et de manioc.

Les structures agricoles de la région ont été influencées par la rigueur du climat et la pauvreté des sols, mais aussi par les conséquences qu'a entraîné la colonisation, notamment en implantant la monoculture arachidière au détriment de la polyculture traditionnelle de subsistance et en modifiant le mode d'appropriation du sol qui était plutôt collectif pendant la période pré-coloniale. Aujourd'hui, dans ce village de 496 habitants vivant pauvrement d'agriculture et d'élevage de subsistance, les problèmes reliés à l'exhaure de l'eau, à la cuisson des aliments et à l'éclairage sont cruciaux.

Les solutions proposées afin de résoudre ces problèmes sont les suivantes:

- utilisation d'une éolienne servant à l'exhaure de l'eau
- implantation d'un système de cuisinière solaire pour compenser la rareté du bois utilisé actuellement pour la cuisson des aliments qui est une activité très pénible
- recyclage de la fumure animale en biogaz (méthane) devant servir à produire l'électricité nécessaire à l'éclairage.

Cette recherche se veut être avant tout une étude de faisabilité afin de démontrer qu'il est possible et viable d'utiliser les énergies renouvelables, éolienne, solaire et du biogaz afin de remplacer des techniques et des énergies traditionnelles qui n'ont peut-être dans certains cas, plus leur place, soit pour des raisons économiques, environnementales ou sociales.

L'utilisation de ces nouvelles techniques énergétiques s'inscrit dans un vaste programme des Nations-Unies pour l'environnement mettant de l'avant l'exploitation des énergies renouvelables au sein des régions rurales des pays en voie de développement. Le but à long terme, est de permettre de généraliser l'application de cette étude à toutes les régions du Sahel.

La méthode générale s'inspire peu d'une étude d'impact environnementale. La raison est que la technologie proposée n'entraîne en fait aucune atteinte au milieu naturel sauf peut-être aux habitudes humaines.

Cette recherche vise à améliorer une situation sans compliquer inutilement le processus traditionnels de fonctionnement et corriger les impacts de technologie inadaptées et imposées.

L'énergie solaire étant proposée pour remplacer le bois pour la cuisson, l'énergie éolienne pour remplacer l'effort manuel des villageois pour pomper l'eau des puits et permettre en plus un stockage appréciable, et finalement l'utilisation du fumier étant mis de l'avant comme source d'énergie électrique, et étant donné que les techniques reliées à ces énergies existent et ont fait leurs preuves, on démontre ici que ces énergies sont bien disponibles en quantités suffisantes et à long terme.

Dans un premier temps, une description du milieu biophysique et humain est présentée afin de connaître la réalité et les conditions actuelles du milieu à l'étude pour répondre aux différents besoins de la population.

Dans un deuxième temps, une cueillette et une analyse approfondie est faite de toutes les données existantes reliées à ces trois sources énergétiques, afin d'en démontrer la disponibilité.

Une présentation de l'équipement proposé ainsi qu'un scénario possible suivent ces deux premières parties pour résoudre les problèmes énergétiques.

TABLE DES MATIERES

Remerciements.....	ii
Résumé.....	iii
Table des matières.....	v
Liste des cartes.....	viii
Liste des tableaux.....	ix
Liste des figures.....	xi
Introduction.....	1
I. Aperçu sur les milieux physique et humain.....	3
A. Aperçu général du village de Ndia	3
Gouréyé.....	6
B. Le relief, la géologie et les sols.....	10
C. Le climat et la pluviométrie.....	14
D. Le vent.....	18
E. La radiation solaire.....	29
F. La végétation et le paysage agraire.....	
G. L'évolution historique de la structure	32
agro-économique.....	34
H. La population et l'organisation spatiale.....	
I. Critères de sélection du site de Ndia	38
Gouréyé.....	44
J. Sommaire de la question énergétique.....	
II. Sources d'énergies disponibles pour répondre	
aux besoins en eau, en énergie de cuisson et	49
en éclairage.....	49
A. L'exhaure de l'eau.....	
1. Les problèmes reliés à l'exhaure	49
de l'eau.....	
2. Source d'énergie utilisable:	50
le vent.....	
3. Estimation de la performance des	
aérogénérateurs en fonction des	52
vent disponible au site.....	

B. La cuisson.....	56
1. Les problèmes reliés à la cuisson.....	56
2. Source d'énergie utilisable: la radiation solaire.....	58
a) estimation de l'énergie solaire nécessaire pour la cuisson.....	58
C. L'éclairage et la production d'électricité.....	64
1. Les problèmes reliés à l'éclairage et à la production d'électricité.....	64
2. Source d'énergie utilisable: le biogaz.....	65
a) le biogaz disponible au site.....	65
b) la quantité de biogaz nécessaire.....	67
3. Source d'énergie alternative ou complémentaire potentiellement utilisable: le vent.....	71
D. Performance technique des systèmes énergétiques du village.....	74
III. Recommandations.....	79
A. L'équipement recommandé.....	79
B. Scénario possible pour Ndia Gouréyé.....	81
1. Demande en eau.....	81
2. Demande en énergie de cuisson.....	83
3. Demande en énergie électrique (éclairage).....	85
4. Disposition générale.....	89
C. Recommandations finales.....	90
Conclusion.....	93

Bibliographie.....	96
A. Le milieu physique et humain.....	96
B. L'énergie de la biomasse.....	98
C. L'énergie éolienne.....	99
D. L'énergie solaire.....	100
E. Organismes et personnes ressources au Sénégal.....	100
F. Manufacturiers.....	101
Annexes.....	102
A. Tableaux et graphiques concernant l'énergie éolienne.....	102
B. Schémas concernant l'équipement recommandé.....	113
C. Questionnaire sur les villages visités.....	133
D. Données climatologiques complémentaires.....	139

