

Michèle Hudon, *Le thésaurus : conception, élaboration, gestion*, Québec (Canada) : Éditions ASTED, 1994, 220 p. (Clé en main)

Introduction

L'intérêt pour certaines formes de langages contrôlés d'indexation et de recherche n'a cessé de croître au cours de la dernière décennie. Cet intérêt renouvelé est dû, en partie au moins, à la prolifération des centres de documentation et d'information spécialisés de même qu'à la disponibilité élargie des logiciels de gestion de bases de données grâce auxquels ces centres peuvent créer leur propre système d'accès aux documents conservés sur place. L'accès or sujet est primordial. Il doit être sophistiqué dans ces milieux où la majorité des documents appartiennent à des domaines connexes et traitent de sujets voisins. L'indexation et la recherche par mots-clés sont possibles et généralement efficaces si l'utilisateur pose une question bien précise et se satisfait d'une ou deux références pertinentes. L'indexation et la recherche au moyen de descripteurs contrôlés sont avantageuses dans les milieux où les questions de l'utilisateur sont générales et où le taux de rappel est de première importance.

Le thésaurus de descripteurs s'est imposé au cours de cette période comme l'outil d'analyse et de recherche documentaire le plus fréquemment utilisé dans les grandes bases de données américaines et européennes.

Les centres de documentation et d'information spécialisés ont accès aux grands thésaurus (*Macrothésaurus pour le traitement de l'information relative au développement économique et social*, *Thésaurus européen de l'éducation*, *Thésaurus of ERIC Descriptors*, *NASA Thésaurus*, etc.), des outils dont ils font d'ailleurs un usage constant. Il arrive cependant que se manifeste le besoin de descripteurs très spécifiques dans un domaine du savoir couvert ou non par un grand thésaurus. Les centres de documentation et d'information se concentrent en effet sur des champs de recherche et d'activités très spécialisés qui ne sont que marginalement couverts par les thésaurus existant dans des disciplines connexes, et les descripteurs spécifiques dont leurs utilisateurs ont besoin ne se retrouvent pas dans les thésaurus qui leur sont disponibles.

Le développement et la mise en disponibilité d'un thésaurus original dans ces champs de recherche en ex-

pansion permettent une organisation plus efficace et un repérage de meilleure qualité dans les collections de documents qui portent sur des sujets nouveaux ou qui offrent de nouvelles perspectives sur des thèmes connus. Ils facilitent également la communication et l'échange d'information entre institutions et entre spécialistes intéressés à ces disciplines.

La grande majorité des thésaurus qui existent déjà ou qui sont en développement ont une structure similaire qui les rend compatibles et qui facilite une utilisation optimale par l'indexeur et par le chercheur. La construction d'un thésaurus se fait à partir d'un ensemble de règles énoncées dans des cahiers de normes internationales et nationales. Ces règles et directives ne sont cependant pas toujours suffisantes pour résoudre les problèmes qui se présentent quotidiennement lors d'un projet de création de thésaurus.

Le présent ouvrage s'adresse au spécialiste de l'information et de la documentation

- qui s'intéresse au thésaurus, à son contenu et à sa structure;
- qui s'interroge sur la nécessité de développer un nouveau thésaurus;
- qui a déjà pris la décision de développer un nouveau thésaurus
- qui a entrepris le développement d'un thésaurus et doit maintenant faire face à des difficultés prévues ou imprévues.

Nous y adoptons une approche résolument pratique tout en mettant l'accent sur les tâches intellectuelles plutôt que sur l'aspect technologique des opérations. Sans ignorer les bases théoriques de la construction des thésaurus, nous discutons de l'organisation des tâches, des variables à considérer, des décisions à prendre, et des solutions possibles à une sélection de problèmes communément rencontrés. Les nombreux exemples qui parsèment le texte devraient en faciliter la lecture et permettre une meilleure compréhension des problèmes et des solutions proposées.

Le thésaurus que nous décrivons ici est le thésaurus « traditionnel » tel que nous le connaissons dans le monde occidental, c'est-à-dire le thésaurus à termes préférentiels établi a priori, soit en relation avec une base de données particulière, soit encore comme outil de référence indépendant servant de base terminologique ou de base de connaissances.

1/ Le thésaurus de descripteurs

1.1/ Un peu d'histoire

C'est à partir de 1947 que l'on retrouve dans la littérature des sciences de l'information des références fréquentes à la notion de thésaurus de descripteurs. Le premier thésaurus opérationnel paraît en 1959. Le contenu, le mode de présentation et la fonction du thésaurus sont fermement établies au début des années 1970. Les années 1980 sont désormais considérées comme la décennie du thésaurus.

1.2/ Qu'est-ce qu'un thésaurus?

1.2.1/ Définition

1.2.2/ Lexique et structure

1.2.2.1/ *Lexique*

1.2.2.2/ *Définitions et notes d'application*

1.2.2.3/ *Structure classificatoire*

1.2.3/ Présentation du contenu des thésaurus

1.2.3.1/ *Présentation alphabétique complète*

1.2.3.2/ *Présentation alphabétique des mots-clés*

1.2.3.3/ *Présentation hiérarchique*

1.2.3.4/ *Présentation thématique*

1.2.3.5/ *Présentation graphique bidimensionnelle*

1.2.4/ Rôle et fonction

Le thésaurus se définit d'abord par sa nature : liste structurée de mots simples ou d'expressions. On le définit ensuite par sa fonction : utilisable pour indexer le contenu des documents et des questions dans un système de transfert d'information.

Tout thésaurus est constitué d'un lexique comprenant des descripteurs et des non-descripteurs, d'un ensemble de définitions et de notes d'application et d'une structure classificatoire s'exprimant par des relations sémantiques d'équivalence, de hiérarchie et d'association.

1.3/ Thésaurus et autres sources terminologiques

1.3.1/ Thésaurus de descripteurs et dictionnaire

1.3.2/ Thésaurus de descripteurs et thésaurus littéraire

1.3.3/ Thésaurus de descripteurs et banque de terminologie

1.3.4/ Thésaurus de descripteurs et liste de vedettes-matière

Il existe cinq types traditionnels de présentation du contenu d'un thésaurus : la présentation alphabétique complète, la présentation alphabétique des mots-clés, la présentation hiérarchique, la présentation thématique et la présentation graphique bidimensionnelle.

<u>2/ La structure sémantique du thésaurus</u>	Les relations sémantiques jouent un rôle important au niveau de la définition des descripteurs puisqu'elles servent à le situer précisément dans l'univers sémantique décrit.
<i>2.1/ Équivalence intralinguistique</i> 2.1.1/ Définition et fonction 2.1.2/ Typologie 2.1.3/ Notation 2.1.4/ Établissement des relations d'équivalence intralinguistique <i>2.2/ Hiérarchie</i> 2.2.1/ Définition et fonction 2.2.2/ Typologie 2.2.3/ Notation 2.2.4/ Établissement des relations hiérarchiques <i>2.3/ Association</i> 2.3.1/ Définition et fonction 2.3.2/ Typologie 2.3.3/ Notation 2.3.4/ Établissement des relations associatives <i>2.4/ Équivalence interlinguistique</i> 2.4.1/ Définition et fonction 2.4.2/ Typologie 2.4.3/ Notation 2.4.4/ Établissement des relations d'équivalence interlinguistique	L'équivalence intralinguistique lie tous les mots ou expressions d'une même langue qui expriment la même idée ou désignent le même objet. La relation hiérarchique exprime un rapport de supériorité ou de subordination entre notions. La relation de type générique lie deux descripteurs dans un rapport de genre à espèce. La relation de type partitif lie deux descripteurs dans un rapport de tout à partie. La relation associative lie des concepts souvent associés mentalement par les spécialistes d'un domaine. Le lien associatif est cependant mal défini et très subjectif. L'équivalence interlinguistique lie les termes de langues différentes qui représentent le même concept ou un concept similaire. On distingue plusieurs degrés d'équivalence interlinguistique, allant de l'équivalence complète à la mono-équivalence.

3/ La décision de créer un nouveau thésaurus

3.1/ Niveaux de contrôle du vocabulaire

Le développement d'un thésaurus est coûteux en temps, en énergie humaine et en argent, et la décision de développer un nouveau thésaurus doit toujours être précédée d'une analyse systématique des besoins et des outils existants.

3.2/ Vocabulaires d'indexation et de recherche déjà disponibles

3.2.1/ Identification des vocabulaires disponibles

3.2.2/ Évaluation des vocabulaires disponibles

L'évaluation des langages documentaires déjà disponibles se fait sur la base des critères suivants : la richesse conceptuelle, la richesse lexicale, l'organisation sémantique.

Il faut éviter de « réinventer la roue » : si un thésaurus approprié existe, il doit être adopté.

3.3/ Publication de la décision de développer un nouveau thésaurus

Il faut publiciser la décision de développer un nouveau thésaurus et encourager la participation de toutes les parties intéressées.

4/ L'analyse des besoins particuliers et la définition des spécifications

4.1/ Analyse des besoins particuliers à l'environnement

- 4.1.1/ Domaine(s) à couvrir**
- 4.1.2/ Types de documents à traiter**
- 4.1.3/ Nombre de documents à traiter**
- 4.1.4/ Personnalité de l'utilisateur principal**
- 4.1.5/ Besoins de l'utilisateur principal : les questions**
- 4.1.6/ Fréquence d'utilisation du système**
- 4.1.7/ Index**

4.2/ Évaluation des ressources

4.3/ Sélection des caractéristiques essentielles du thésaurus

- 4.3.1/ Limites du thésaurus (domaines/-disciplines/ sujets)**
- 4.3.2/ Profondeur du thésaurus (volume/-nombre de descripteurs)**
- 4.3.3/ Structure théaurale**
- 4.3.4/ Présentation du produit final**
- 4.3.5/ Traitement des données**

L'analyse des besoins particuliers à l'environnement constitue la première tâche à accomplir dans l'élaboration du thésaurus. L'analyse des besoins vise notamment à caractériser les domaines ou disciplines qui seront décrits dans le thésaurus, à identifier les types de documents à traiter et à en estimer le nombre, à étudier la personnalité et les besoins de l'utilisateur principal.

Il faut également prendre en début de projet des décisions très importantes sur les limites du thésaurus (définition des domaines qui seront couverts), la profondeur du thésaurus (nombre de descripteurs), la complexité de la structure théaurale, la présentation du produit final et le mode de traitement des données.

5/ La construction du thésaurus

5.1/ Normes et directives

Le respect des principes directeurs nationaux et internationaux pour le développement des thésaurus assurent à un thésaurus une cohérence interne et un degré élevé de compatibilité avec les outils du même type.

5.2/ Collecte des termes

5.2.1/ Approche analytique

5.2.2/ Approche synthétique

5.2.3/ La collecte ... en pratique

5.2.3.1/ *Combinaison des deux approches*

5.2.3.2/ *Recherche et sélection des sources*

5.2.3.3/ *Identification des candidats-descripteurs*

5.2.3.4/ *Enregistrement des candidats-descripteurs*

5.2.3.5/ *Durée de la collecte*

La combinaison des deux approches reconnues pour la collecte des termes donne les meilleurs résultats pour ce qui est du contenu et du volume de la banque de candidats-descripteurs. L'approche analytique privilégie la littérature courante et les questions d'utilisateurs comme source de termes. L'approche synthétique se concentre plutôt sur les ouvrages de référence dans les domaines à décrire.

5.3/ Classification des candidats-descripteurs / Premiers groupements sémantiques

5.3.1/ Regroupement par thèmes

5.3.2/ Regroupement par facettes

5.3.3/ Combinaison des deux méthodes de regroupement

On effectue une première analyse du contenu de la banque de termes pour regrouper les concepts reliés et faciliter les étapes subséquentes de sélection des descripteurs et d'établissement des relations sémantiques. On peut grouper les concepts et les termes selon leur appartenance à un grand thème ou sur la base de facettes pertinentes aux domaines décrits.

5.4/ Choix des descripteurs

5.4.1/ Identification des équivalents conceptuels

5.4.1.1/ *Synonymie véritable*

5.4.1.2/ *Quasi-synonymie*

5.4.1.3/ *Antonymie*

5.4.2/ Sélection des descripteurs

5.4.2.1/ *Critères de sélection*

5.4.2.2/ *Forme des descripteurs*

5.4.2.3/ *Précoordination des concepts et des termes*

5.4.2.4/ *Homographes*

5.4.2.5/ *Définitions et notes d'application*

5.4.2.6/ *Noms propres*

5.4.2.7/ *Mots « à la mode »*

5.5/ Reconstitution et présentation des hiérarchies

5.5.1/ Validation des relations génériques

5.5.2/ Polyhiérarchie

5.5.3/ Les «orphelins»

5.5.4/ Niveaux hiérarchiques

5.5.5/ Présentation des hiérarchies

5.5.5.1/ *Présentation des termes spécifiques*

5.5.5.2/ *Relation générique et relation partitive*

Dans l'ensemble des termes qui décrivent une même notion (synonymes, quasi-synonymes et antonymes), le descripteur est idéalement celui qui est le plus connu et le plus fréquemment utilisé dans les documents à traiter et dans les questions des usagers. Les termes qui ne sont pas choisis comme descripteurs deviennent des non-descripteurs.

Les principes directeurs pour le développement des thésaurus sont explicites en ce qui concerne la nature et la forme des descripteurs. Le descripteur est un nom ou une expression nominale exprimant un seul concept et généralement présenté au masculin et au singulier.

Il existe plusieurs façons de s'assurer de la validité d'une relation hiérarchique de type générique où le concept représenté par le terme spécifique est toujours et nécessairement un type particulier du concept représenté par le terme générique.

Au sein des domaines principaux qui sont décrits par le thésaurus, les hiérarchies sont plus développées et comprennent plusieurs niveaux. La présentation alphabétique favorisée par la majorité des thésaurus ne révèle cependant sous chaque descripteur qu'un seul niveau supérieur et un niveau inférieur.

5.6/ Établissement et présentation des relations associatives

5.6.1/ Types d'associations

5.6.2/ Validité des associations

5.6.3/ Nombre de termes associés

Bien qu'il n'existe pas de règles strictes pour l'établissement de relations associatives entre descripteurs, on peut établir une liste de types ou de catégories de concepts qu'il est utile d'associer. Le nombre de termes associés à un descripteur doit être restreint si on veut que l'association joue bien son rôle, celui de contribuer à la définition du descripteur.

5.7/ Enregistrement et manipulation des données

5.7.1/ Organisation générale du travail

5.7.2/ Division du travail

5.7.3/ Le dossier descripteur

5.7.4/ Enregistrement des données

5.7.4.1/ Le support « papier »

5.7.4.2/ Le support informatique

5.7.5/ Utilité des représentations graphiques

5.7.6/ Listes de travail

5.7.7/ Utilisation du thésaurus en développement pour l'indexation

Dans un projet de construction de thésaurus, l'organisme responsable doit développer des procédures et des méthodes de travail en tenant compte des ressources disponibles et des préférences du personnel impliqué. La définition et l'établissement d'un dossier descripteur est essentielle. Le mode d'enregistrement des données, l'utilisation des supports « papier » et informatique et l'application du thésaurus en développement à l'indexation des documents courants font l'objet de décisions prises localement.

5.8/ Révision et vérification des données thésaurales

5.9/ Le produit final

5.9.1/ Quand peut-on «produire» un thésaurus?

5.9.2/ Version imprimée

5.9.2.1/ *Les diverses sections d'un thésaurus*

5.9.2.2/ *Présentation et disposition des données*

5.9.2.3/ *Révision et vérification du document imprimé*

5.9.3/ Version électronique

5.9.3.1/ *Types d'utilisateurs*

5.9.3.2/ *Types de présentation*

5.9.4/ Dépôt et mise en disponibilité du document

Des vérifications formelles et structurales doivent être effectuées pour s'assurer de la qualité et de l'intégrité du contenu d'un thésaurus. L'apport de l'informatique se révèle ici inestimable.

On doit considérer qu'un thésaurus est fonctionnel lorsqu'on estime qu'il contient au moins 80% des concepts et des termes qui décrivent les domaines couverts.

La plupart des thésaurus sont encore présentés sous forme imprimée et constituent un document comprenant plusieurs sections. La partie principale est celle qui donne le plus d'information sur chaque terme et qui reprend la plus grande partie du dossier descripteur. Les présentations en ligne de ces mêmes thésaurus sont encore souvent une transposition sur écran de la page imprimée.

Lorsque le thésaurus est finalement disponible, il faut en avvertir toutes les parties qui étaient déjà au courant de son développement et déposer le document dans des collections de référence appropriées.

<u>6/ La validation du contenu et de la pertinence</u>	On peut et on doit constamment contrôler la qualité du contenu d'un thésaurus en développement.
<i>6.1/ Sources documentaires</i> <i>6.2/ Personnes-ressources</i> 6.2.1/ Rôle et responsabilité du comité consultatif 6.2.2/ Composition du comité consultatif 6.2.3/ Modalités de participation <i>6.3 Mise à l'essai</i> 6.3.1/ Objectifs de la mise à l'essai 6.3.2/ Types de mise à l'essai 6.3.3/ Responsabilité de la mise à l'essai 6.3.4/ Les opérations 6.3.5/ Résultats attendus	Les sources utilisées pour la collecte des candidats-descripteurs et l'établissement d'une structure classificatoire doivent être représentatives des domaines à décrire et être reconnues comme étant complètes et de grande qualité. Il est recommandé qu'un comité consultatif de spécialistes choisis en fonction de leur connaissance pratique de l'organisation conceptuelle et de la terminologie des domaines à décrire soit disponible pour valider les décisions prises par l'équipe de travail et pour suggérer des solutions aux cas problèmes. La mise à l'essai d'une version préliminaire du thésaurus sur un échantillon représentatif de documents et de questions permet de confirmer que le lexique et la structure déjà établis répondent aux besoins, tout en permettant d'identifier les descripteurs, non-descripteurs et relations sémantiques qui manquent encore.

7/ La gestion du projet

7.1/ Personnel

Il est préférable de faire appel à des spécialistes de la construction des thésaurus pour mener à bien un projet de développement de thésaurus.

Si une institution confie la tâche d'élaborer un thésaurus à des membres de son personnel, il est préférable que ceux-ci soient dégagés de toute autre responsabilité et qu'ils se voient offrir d'abord une formation de base dans ce domaine.

7.2/ Échéancier

Il faut établir dès le départ un échéancier précis indiquant les dates de début et de fin de chacune des grandes étapes du développement d'un thésaurus.

7.3/ Coûts

Le coût total du développement d'un thésaurus varie en fonction de facteurs liés au thésaurus lui-même (limites, volume, etc.) ainsi que de facteurs liés à l'équipe responsable du développement (expérience, disponibilité, etc.). Le développement d'un thésaurus est toujours coûteux.

7.4/ Bureaucratie

8/ La mise à jour

8.1/ Objectifs de la mise à jour

Le développement d'un thésaurus n'est jamais terminé. Un thésaurus doit être mis à jour régulièrement pour continuer à refléter l'évolution des connaissances dans un domaine ainsi que les intérêts et les besoins des usagers.

8.2/ Nécessité de la mise à jour

8.2.1/ Lacunes conceptuelles et terminologiques

8.2.2/ Fréquence d'utilisation des descripteurs

8.2.3/ Difficulté d'utilisation du thésaurus

La mise à jour d'un thésaurus permet de corriger les erreurs et de combler les lacunes conceptuelles et terminologiques qui y subsistent.

8.3 /Les opérations

8.3.1/ Ajout de nouveaux termes

8.3.2/ Modification de descripteurs existants

8.3.2.1/ Modification formelle simple du descripteur

8.3.2.2/ Modification de l'énoncé du descripteur

8.3.2.3/ Ajout de liens sémantiques

8.3.2.4/ Explication du contexte d'utilisation

8.3.3/ Suppression de descripteurs

Lors de la mise à jour, les opérations suivantes sont effectuées : ajout de nouveaux termes et création de nouveaux dossiers descripteurs, modification de descripteurs et de dossiers existants, suppression de termes et de dossiers existants.

Il faut toujours agir avec prudence : toute modification au contenu original d'un thésaurus peut entraîner une cascade de changements structuraux dans le thésaurus même ainsi que dans les index pour lesquels il sert d'instrument de contrôle du vocabulaire.

8.4/ Gestion de la mise à jour

8.4.1/ Responsabilité générale

8.4.2/ Collecte des données

8.4.3/ Fréquence de mise à jour

8.4.4/ Contrôle de la qualité

8.4.5/ Présentation des ajouts et des modifications

Il est préférable que l'institution et les personnes responsables de la construction du thésaurus se chargent également de sa mise à jour. Les utilisateurs du thésaurus doivent être invités à participer activement aux travaux de mise à jour. Le rythme d'expansion et la qualité des ajouts et des modifications doivent cependant être contrôlés en tout temps.

<p><u>9/ Un cas particulier : le thésaurus multilingue</u></p>	<p>Le thésaurus multilingue offre un inventaire complet des concepts, des termes et de leurs relations dans chacune des langues décrites.</p>
<p><i>9.1/ Construction du thésaurus multilingue</i></p> <p>9.1.1/ Diverses approches</p> <p>9.1.2/ Analyse environnementale préliminaire</p> <p>9.1.3/ Collecte des termes</p> <p>9.1.4/ Identification des équivalents interlinguistiques</p> <p>9.1.4.1/ <i>Équivalence exacte</i></p> <p>9.1.4.2/ <i>Équivalence inexacte</i></p> <p>9.1.4.3/ <i>Équivalence partielle</i></p> <p>9.1.4.4/ <i>Équivalence simple à multiple</i></p> <p>9.1.4.5/ <i>Non-équivalence</i></p> <p>9.1.5/ Sélection des descripteurs</p> <p>9.1.6/ Forme des descripteurs</p> <p>9.1.7/ Hiérarchies et associations</p> <p>9.1.7.1/ <i>Relations hiérarchiques</i></p> <p>9.1.7.2/ <i>Relations associatives</i></p> <p>9.1.8/ Vérification et validation</p> <p>9.1.9/ Présentation</p> <p>9.1.10/ Mise à jour</p> <p><i>9.2/ Gestion du projet</i></p> <p>9.2.1/ Structure administrative</p> <p>9.2.2/ Personnel</p> <p>9.2.3/ Échéancier</p> <p>9.2.4/ Coûts</p>	<p>L'élaboration simultanée des diverses versions linguistiques d'un thésaurus est la seule approche qui garantisse l'égalité des langues et des cultures décrites. La traduction d'un thésaurus unilingue déjà disponible n'est pas recommandée puisqu'elle donne à la langue source un statut dominant et qu'elle encourage l'impérialisme culturel.</p> <p>Il existe divers niveaux d'équivalence entre des termes appartenant à des langues différentes. L'équivalence exacte, l'équivalence inexacte et l'équivalence partielle ne posent pas de problèmes au niveau de l'élaboration du thésaurus multilingue. L'équivalence simple à multiple et la non-équivalence présentent parfois des problèmes sérieux pour lesquels on dispose cependant d'une variété de solutions acceptables.</p> <p>La gestion semi-centralisée d'un projet d'élaboration de thésaurus multilingue semble la plus efficace et la plus satisfaisante pour toutes les parties intéressées.</p> <p>La production d'un thésaurus bilingue demandera probablement quelques mois de plus et coûtera environ une fois et demie plus cher que celle du thésaurus unilingue.</p>

Glossaire

BRUIT	Toute réponse non pertinente obtenue lors d'une recherche documentaire. (<i>voir aussi</i> Taux de précision.)
CANDIDAT-DESCRIPTEUR	Terme retenu pour insertion éventuelle dans un thésaurus.
COMPATIBILITÉ	En général, caractère de ce qui peut s'accorder, se concilier, fonctionner ensemble. En contexte, caractère de plusieurs thésaurus établis selon les mêmes méthodes et dont les descripteurs peuvent s'additionner et être employés pour indexer ou rechercher des documents dans un même domaine.
CONCEPT	(ou <i>notion</i>) Élément de la pensée, représentation mentale d'êtres ou de choses de qualités, d'actions, de localisations, de situations, de rapports, etc. le plus souvent exprimés par un terme.
DESCRIPTEUR	Dans un thésaurus, terme retenu et choisi parmi un ensemble de termes synonymes pour représenter sans ambiguïté un concept contenu dans un document ou dans une stratégie de recherche documentaire. (<i>voir aussi</i> Non-descripteur, Thésaurus à termes préférentiels.)
ÉQUIVALENCE INTERLINGUISTIQUE	voir Équivalence linguistique.
ÉQUIVALENCE INTRALINGUISTIQUE	(relation d'équivalence <i>ou</i> relation préférentielle) Dans un thésaurus, lien qui existe entre les divers termes utilisés pour représenter le même concept.
ÉQUIVALENCE LINGUISTIQUE	(ou <i>équivalence interlinguistique</i>) Dans un thésaurus multilingue, lien qui existe entre les termes de langues différentes qui représentent le même concept. (<i>voir aussi</i> Thésaurus multilingue.)
FACETTE	Catégorie de notions de même nature tels que processus, matériau, propriété, etc. permettant un regroupement de termes en dehors du contexte d'une discipline.
HOMOGRAPHE	Terme dont la forme graphique est identique à celle d'un autre terme mais qui diffère de celui-ci par l'origine et par le sens.
IDENTIFICATEUR	Nom propre représentant un concept unique (individu, projet, organisme, etc.) utilisable pour l'indexation et la recherche documentaire.

LANGAGE CONTRÔLÉ	<i>voir</i> Langage documentaire.
LANGAGE D'INDEXATION	<i>voir</i> Langage documentaire.
LANGAGE DOCUMENTAIRE	(langage d'indexation <i>ou</i> langage contrôlé) Langage artificiel constitué de représentations de concepts et de relations entre ces concepts, et destiné dans un système documentaire à formaliser les données contenues dans les documents sources et dans les demandes des utilisateurs.
MONOHIÉRARCHIE	Caractéristique d'un langage documentaire dans lequel chaque terme ne peut appartenir à plus d'une hiérarchie et ne peut avoir qu'un terme générique de niveau immédiatement supérieur.
MOT-CLÉ	Mot ou expression du langage naturel utilisé pour l'indexation ou la recherche documentaire.
NON-DESCRIPTEUR	Dans un thésaurus, terme qui appartient au lexique mais qui ne peut être utilisé pour l'indexation ou la recherche Documentaire. Le non-descripteur est toujours lié au descripteur choisi pour représenter le même concept. (<i>voir aussi</i> Descripteur; Thésaurus à termes préférentiels)
NOTE D'APPLICATION	Dans un thésaurus, brève explication ou directives pour l'emploi d'un descripteur.
NOTION	<i>voir</i> Concept.
POLYHIÉRARCHIE	Caractéristique d'un langage documentaire dans lequel chaque terme peut appartenir à plusieurs hiérarchies et peut avoir plus d'un terme générique de niveau immédiatement supérieur.
POSTCOORDINATION	Dans un système documentaire, principe suivant lequel l'expression d'un sujet s'effectue au moment de la recherche documentaire par combinaison de plusieurs termes ou descripteurs.
PRÉCOORDINATION	Dans un système documentaire, principe suivant lequel l'expression d'un sujet s'effectue au moment de l'élaboration du vocabulaire d'indexation ou au moment de l'indexation par juxtaposition de termes et création de descripteurs composés, de vedettes-matière, de chaînes de descripteurs, etc.
QUASI-SYNONYME	(<i>ou</i> synonyme documentaire) Terme dont le sens n'est pas exactement le même que celui d'un autre terme mais qui dans un certain contexte est considéré comme son équivalent.

RELATION ASSOCIATIVE	Dans un thésaurus, lien établi entre des descripteurs représentant des concepts que les usagers associent mentalement.
RELATION D'ÉQUIVALENCE	<i>voir Équivalence intralinguistique.</i>
RELATION HIÉRARCHIQUE	Dans un thésaurus, lien entre deux descripteurs dont l'un est subordonné à l'autre.
RELATION HIÉRARCHIQUE GÉNÉRIQUE	Relation hiérarchique au sein de laquelle le concept représenté par le terme spécifique est un type ou une espèce particulière du concept représenté par le terme générique.
RELATION HIÉRARCHIQUE PARTITIVE	Relation hiérarchique au sein de laquelle le concept représenté par le terme spécifique constitue une partie du tout décrit par le terme générique.
RELATION PRÉFÉRENTIELLE	<i>voir Équivalence intralinguistique.</i>
SILENCE	Ensemble des documents pertinents qui ne sont pas repérés lors d'une recherche documentaire. (<i>voir aussi Taux de rappel</i>)
SYNONYME DOCUMENTAIRE	<i>voir Quasi-synonyme.</i>
SYNONYMES	Termes qui ont le même sens mais des formes graphiques différentes. (<i>voir aussi Équivalence intralinguistique</i>)
TAUX DE PRÉCISION	Nombre de documents pertinents repérés lors d'une recherche documentaire par rapport au nombre total de documents obtenus en réponse à la question posée. (<i>voir aussi Bruit.</i>)
TAUX DE RAPPEL	Nombre de documents pertinents repérés lors d'une recherche documentaire par rapport au nombre total de documents pertinents que l'on estime être présents dans la collection. (<i>voir aussi Silence</i>)
TERME	Mot ou groupe de mots représentant un concept.
THÉSAURUS	Langage documentaire fondé sur une structuration hiérarchisée d'un ou plusieurs domaines de la connaissance et dans lequel les notions sont représentées par des termes de la langue naturelle et les relations entre notions par des signes conventionnels. (AFNOR 1987)

**THÉSAURUS À
TERMES NON
PRÉFÉRENTIELS**

Thésaurus dans lequel tous les termes représentant un concept sont utilisables pour l'indexation et pour la recherche documentaire.

**THÉSAURUS À
TERMES
PRÉFÉRENTIELS**

Thésaurus dans lequel un seul des termes représentant un concept est utilisable pour l'indexation et la recherche documentaire. (*voir aussi* **Descripteur, Non-descripteur**)

**THÉSAURUS
MONOLINGUE**

Thésaurus dont tous les descripteurs appartiennent à une seule langue.

**THÉSAURUS
MULTILINGUE**

Thésaurus contenant des descripteurs appartenant à plus d'une langue. (*voir aussi* **Équivalence linguistique**)

Conclusion

Le thésaurus est un outil de travail dont l'importance et l'utilité ne sont plus mises en doute. La disponibilité d'un thésaurus complet, logiquement structuré et régulièrement mis à jour pour intégrer de nouveaux concepts et répondre à de nouveaux besoins ne peut que contribuer à l'amélioration marquée des résultats de l'indexation et du repérage dans un système d'information automatisé ou non.

Il ne faut pas oublier cependant que le thésaurus n'est qu'une des composantes d'un tel système. L'instrument de contrôle du vocabulaire que nous avons développé ne sera pleinement efficace que s'il est utilisé dans le cadre de l'application d'une politique d'indexation bien claire. Le thésaurus n'est qu'un élément parmi d'autres dans le système d'indexation. Il ne suffit pas de disposer d'un langage d'indexation : il faut savoir d'abord ce qu'on indexe, pour qui et pourquoi on indexe, et finalement comment on indexe.

Rappelons enfin que la disponibilité d'un thésaurus, même d'un thésaurus d'excellente qualité, ne suffit pas non plus à transformer un individu en indexeur hors pair ou en chercheur expert. La formation de toutes les catégories d'utilisateurs reste essentielle à l'exploitation efficace d'un thésaurus, à l'étape de l'indexation comme à celle du repérage.

Table des matières

INTRODUCTION	1
1/ LE THESAURUS DE DESCRIPTEURS	3
1.1/ Un peu d'histoire	3
1.2/ Qu'est-ce qu'un thésaurus?.....	3
1.2.1/ Définition	3
1.2.2/ Lexique et structure	3
1.2.2.1/ Lexique.....	3
1.2.2.2/ Définitions et notes d'application.....	3
1.2.2.3/ Structure classificatoire.....	3
1.2.3/ Présentation du contenu des thésaurus.....	3
1.2.3.1/ Présentation alphabétique complète	3
1.2.3.2/ Présentation alphabétique des mots-clés.....	3
1.2.3.3/ Présentation hiérarchique	3
1.2.3.4/ Présentation thématique	3
1.2.3.5/ Présentation graphique bidimensionnelle	3
1.2.4/ Rôle et fonction.....	3
1.3/ Thésaurus et autres sources terminologiques.....	3
1.3.1/ Thésaurus de descripteurs et dictionnaire	3
1.3.2/ Thésaurus de descripteurs et thésaurus littéraire.....	3
1.3.3/ Thésaurus de descripteurs et banque de terminologie.....	3
1.3.4/ Thésaurus de descripteurs et liste de vedettes-matière.....	3
2/ LA STRUCTURE SEMANTIQUE DU THESAURUS	4
2.1/ Équivalence intralinguistique	4
2.1.1/ Définition et fonction	4
2.1.2/ Typologie	4
2.1.3/ Notation.....	4
2.1.4/ Établissement des relations d'équivalence intralinguistique.....	4
2.2/ Hiérarchie	4
2.2.1/ Définition et fonction	4
2.2.2/ Typologie.....	4
2.2.3/ Notation	4
2.2.4/ Établissement des relations hiérarchiques.....	4
2.3/ Association	4
2.3.1/ Définition et fonction.....	4
2.3.2/ Typologie.....	4

2.3.3/ Notation.....	4
2.3.4/ Établissement des relations associatives	4
2.4/ Équivalence interlinguistique	4
2.4.1/ Définition et fonction	4
2.4.2/ Typologie.....	4
2.4.3/ Notation.....	4
2.4.4/ Établissement des relations d'équivalence interlinguistique.....	4
3/ LA DECISION DE CREER UN NOUVEAU THESAURUS.....	5
3.1/ Niveaux de contrôle du vocabulaire	5
3.2/ Vocabulaires d'indexation et de recherche déjà disponibles	5
3.2.1/ Identification des vocabulaires disponibles.....	5
3.2.2/ Évaluation des vocabulaires disponibles	5
3.3/ Publication de la décision de développer un nouveau thésaurus	5
4/ L'ANALYSE DES BESOINS PARTICULIERS ET LA DEFINITION DES SPECIFICATIONS	6
4.1/ Analyse des besoins particuliers à l'environnement.....	6
4.1.1/ Domaine(s) à couvrir	6
4.1.2/ Types de documents à traiter.....	6
4.1.3/ Nombre de documents à traiter.....	6
4.1.4/ Personnalité de l'utilisateur principal.....	6
4.1.5/ Besoins de l'utilisateur principal : les questions.....	6
4.1.6/ Fréquence d'utilisation du système	6
4.1.7/ Index	6
4.2/ Évaluation des ressources	6
4.3/ Sélection des caractéristiques essentielles du thésaurus	6
4.3.1/ Limites du thésaurus (domaines/disciplines/ sujets)	6
4.3.2/ Profondeur du thésaurus (volume/nombre de descripteurs).....	6
4.3.3/ Structure théaurale.....	6
4.3.4/ Présentation du produit final	6
4.3.5/ Traitement des données.....	6
5/ LA CONSTRUCTION DU THESAURUS.....	7
5.1/ Normes et directives	7
5.2/ Collecte des termes	7
5.2.1/ Approche analytique	7

5.2.2/ Approche synthétique	7
5.2.3/ La collecte ... en pratique.....	7
5.2.3.1/ Combinaison des deux approches	7
5.2.3.2/ Recherche et sélection des sources.....	7
5.2.3.3/ Identification des candidatsdescripteurs.....	7
5.2.3.4/ Enregistrement des candidatsdescripteurs.....	7
5.2.3.5/ Durée de la collecte	7
5.3/ Classification des candidats-descripteurs / Premiers groupements sémantiques	7
5.3.1/ Regroupement par thèmes.....	7
5.3.2/ Regroupement par facettes.....	7
5.3.3/ Combinaison des deux méthodes de regroupement.....	7
5.4/ Choix des descripteurs	8
5.4.1/ Identification des équivalents conceptuels	8
5.4.1.1/ <i>Synonymie véritable</i>	8
5.4.1.2/ <i>Quasi-synonymie</i>	8
5.4.1.3/ <i>Antonymie</i>	8
5.4.2/ Sélection des descripteurs	8
5.4.2.1/ <i>Critères de sélection</i>	8
5.4.2.2/ <i>Forme des descripteurs</i>	8
5.4.2.3/ <i>Précoordination des concepts et des termes</i>	8
5.4.2.4/ <i>Homographes</i>	8
5.4.2.5/ <i>Définitions et notes d'application</i>	8
5.4.2.6/ <i>Noms propres</i>	8
5.4.2.7/ <i>Mots « à la mode »</i>	8
5.5/ Reconstitution et présentation des hiérarchies	8
5.5.1/ Validation des relations génériques.....	8
5.5.2/ Polyhiérarchie	8
5.5.3/ Les «orphelins».....	8
5.5.4/ Niveaux hiérarchiques	8
5.5.5/ Présentation des hiérarchies	8
5.5.5.1/ <i>Présentation des termes spécifiques</i>	8
5.5.5.2/ <i>Relation générique et relation partitive</i>	8
5.6/ Établissement et présentation des relations associatives	9
5.6.1/ Types d'associations.....	9
5.6.2/ Validité des associations.....	9
5.6.3/ Nombre de termes associés.....	9
5.7/ Enregistrement et manipulation des données.....	9
5.7.1/ Organisation générale du travail	9
5.7.2/ Division du travail	9
5.7.3/ Le dossier descripteur.....	9
5.7.4/ Enregistrement des données.....	9
5.7.4.1/ <i>Le support «papier»</i>	9

5.7.4.2/ Le support informatique	9
5.7.5/ Utilité des représentations graphiques.....	9
5.7.6/ Listes de travail.....	9
5.7.7/ Utilisation du thésaurus en développement pour l'indexation.....	9
5.8/ Révision et vérification des données thésaurales.....	10
5.9/ Le produit final	10
5.9.1/ Quand peut-on «produire» un thésaurus?	10
5.9.2/ Version imprimée.....	10
5.9.2.1/ Les diverses sections d'un thésaurus.....	10
5.9.2.2/ Présentation et disposition des données	10
5.9.2.3/ Révision et vérification du document imprimé.....	10
5.9.3/ Version électronique	10
5.9.3.1/ Types d'usagers	10
5.9.3.2/ Types de présentation	10
5.9.4/ Dépôt et mise en disponibilité du document	10
6/ LA VALIDATION DU CONTENU ET DE LA PERTINENCE	11
6.1/ Sources documentaires	11
6.2/ Personnes-ressources	11
6.2.1/ Rôle et responsabilité du comité consultatif	11
6.2.2/ Composition du comité consultatif.....	11
6.2.3/ Modalités de participation	11
6.3 Mise à l'essai.....	11
6.3.1/ Objectifs de la mise à l'essai.....	11
6.3.2/ Types de mise à l'essai	11
6.3.3/ Responsabilité de la mise à l'essai.....	11
6.3.4/ Les opérations	11
6.3.5/ Résultats attendus	11
7/ LA GESTION DU PROJET.....	12
7.1/ Personnel.....	12
7.2/ Échéancier.....	12
7.3/ Coûts	12
7.4/ Bureaucratie	12
8/ LA MISE A JOUR	13
8.1/ Objectifs de la mise à jour	13
8.2/ Nécessité de la mise à jour	13

8.2.1/ Lacunes conceptuelles et terminologiques	13
8.2.2/ Fréquence d'utilisation des descripteurs.....	13
8.2.3/ Difficulté d'utilisation du thésaurus	13
8.3 /Les opérations	13
8.3.1/ Ajout de nouveaux termes	13
8.3.2/ Modification de descripteurs existants.....	13
8.3.2.1/ <i>Modification formelle simple du descripteur</i>	13
8.3.2.2/ <i>Modification de l'énoncé du descripteur</i>	13
8.3.2.3/ <i>Ajout de liens sémantiques</i>	13
8.3.2.4/ <i>Explication du contexte d'utilisation</i>	13
8.3.3/ Suppression de descripteurs	13
8.4/ Gestion de la mise à jour	13
8.4.1/ Responsabilité générale.....	13
8.4.2/ Collecte des données.....	13
8.4.3/ Fréquence de mise à jour	13
8.4.4/ Contrôle de la qualité	13
8.4.5/ Présentation des ajouts et des modifications.....	13
9/ UN CAS PARTICULIER : LE THESAURUS MULTILINGUE.....	14
9.1/ Construction du thésaurus multilingue	14
9.1.1/ Diverses approches	14
9.1.2/ Analyse environnementale préliminaire.....	14
9.1.3/ Collecte des termes	14
9.1.4/ Identification des équivalents interlinguistiques.....	14
9.1.4.1/ <i>Équivalence exacte</i>	14
9.1.4.2/ <i>Équivalence inexacte</i>	14
9.1.4.3/ <i>Équivalence partielle</i>	14
9.1.4.4/ <i>Équivalence simple à multiple</i>	14
9.1.4.5/ <i>Non-équivalence</i>	14
9.1.5/ Sélection des descripteurs	14
9.1.6/ Forme des descripteurs.....	14
9.1.7/ Hiérarchies et associations.....	14
9.1.7.1/ <i>Relations hiérarchiques</i>	14
9.1.7.2/ <i>Relations associatives</i>	14
9.1.8/ Vérification et validation.....	14
9.1.9/ Présentation.....	14
9.1.10/ Mise à jour	14
9.2/ Gestion du projet.....	14
9.2.1/ Structure administrative	14
9.2.2/ Personnel	14
9.2.3/ Échéancier.....	14

9.2.4/ Coûts	14
GLOSSAIRE	15
CONCLUSION	19
TABLE DES MATIERES.....	20

Lille, le jeudi 12 juin 1997