

**THÉSAURUS DE L'ACTIVITÉ
GOUVERNEMENTALE**

**ORIENTATIONS ET PRINCIPES
DIRECTEURS**

Collection en ingénierie documentaire : 10

Novembre 1998

Réalisé dans le cadre du Chantier en ingénierie documentaire par :
Coordonné par : Richard Parent et Nicole Boulet

Conseil du trésor
Sous-secrétariat à l'information gouvernementale
et aux ressources informationnelles

REMERCIEMENTS

Ce rapport est issu de la réflexion d'un groupe de travail qui s'est penché sur les (objet)

Nous remercions de leur collaboration les personnes suivantes :

Membres du groupe de travail :

Danielle Parent

Claire Fecteau

Réal Dumoulin

Richard Parent

Pierre Noreau

Conseillère : Michèle Hudon (École de bibliothéconomie et des sciences de l'information de l'Université de Montréal)

TABLE DES MATIÈRES

INTRODUCTION	1
CHAPITRE 1	2
LE DOMAINE COUVERT PAR LE PROTOTYPE	2
CHAPITRE 2	3
LE TYPE DE LANGAGE PROPOSÉ POUR LE THÉSAURUS	3
2.1 LANGAGES DOCUMENTAIRES À VOCABULAIRE PRÉ-COORDONNÉ	3
2.2 LANGAGES DOCUMENTAIRES À VOCABULAIRE POST-COORDONNÉ	4
CHAPITRE 3	6
LA STRUCTURE RETENUE POUR LE THÉSAURUS	6
3.1 LA RELATION D'ÉQUIVALENCE.....	6
3.2 LA RELATION HIÉRARCHIQUE	6
3.3 LA RELATION ASSOCIATIVE	7
3.4 LA RELATION DE REGROUPEMENT DE TERMES AU SEIN D'UNE MÊME CATÉGORIE.....	7
3.5 PARTICULARITÉS	12
CHAPITRE 4	13
LES MODALITÉS DE FONCTIONNEMENT À LA DISPOSITION DES INDEXEURS	13
CHAPITRE 5	14
LES MODALITÉS DE FONCTIONNEMENT À LA DISPOSITION DES INTERNAUTES	14
CHAPITRE 6	15
L'ÉLABORATION DU THÉSAURUS	15
ANNEXE 1	1
ANNEXE 2	2

INTRODUCTION

La vague d'Internet et le projet d'infrastructure gouvernementale élargissent la portée de l'information gouvernementale à laquelle l'employé de l'Administration et le citoyen peuvent avoir accès. Un Répertoire gouvernemental est apparu comme un outil essentiel pour favoriser cet accès pour des classes d'informations variées parmi lesquelles se trouvent les données sur les employés (Pages blanches), sur les unités administratives, sur les programmes et services (Pages bleues) ainsi que sur les documents (Pages vertes).

Parmi les moyens de favoriser un repérage sélectif dans cette énorme masse d'information, l'emploi de mots-clés apparaît un élément qualitatif important. Pour que des mots-clés puissent servir à décrire autant des personnes, des programmes gouvernementaux que des documents, un vocabulaire de l'activité gouvernementale s'est imposé. Un tel vocabulaire devrait favoriser un repérage sélectif à l'échelle gouvernementale indépendamment des missions spécialisées et respectives de chacun des ministères et des organismes gouvernementaux.

Par ailleurs, la décentralisation de la création et de la mise à jour des diverses informations du Répertoire gouvernemental amèneront nombre d'employés de l'Administration à saisir celles-ci. L'utilisation d'un vocabulaire contrôlé devient donc un moyen de favoriser l'homogénéité de l'attribution de ces mots-clés.

Ce vocabulaire pourrait n'être qu'un vocabulaire contrôlé, mais l'intensité prévisible de son utilisation tant en mode d'élaboration (employés) qu'en mode de recherche (internautes) justifient qu'il soit structuré dans un support de navigation, fondé sur différents liens entre les mots-clés. Deux options ont été examinées :

Le plan de classification, développé autour de différents sujets et dont l'organisation interne (liens) reflète la collection documentaire à traiter. Il s'agit d'un outil pragmatique.

Le thésaurus, fondé sur une structuration hiérarchisée d'un ou de plusieurs domaines de la connaissance et développé autour de concepts représentés par des termes faisant l'objet des relations de différente nature.

Parce que le thésaurus est plus indépendant des structures gouvernementales et de leur mutation qu'une structure classificatoire par sujets, cet outil a été préféré au plan de classification.

Le présent document énonce les principes qui guideront l'élaboration d'une première version du thésaurus de l'activité gouvernementale. Comme on le verra dans le prochain chapitre, le prototype couvrira, dans un premier temps, la section « Pages bleues » du Répertoire gouvernementale. Les chapitres suivants présentent les types de langage possible pour ce genre d'outil ainsi que les arguments qui ont milité en faveur du thésaurus et de la structure retenue. Enfin les derniers chapitres précisent les différentes modalités qui seront offertes aux utilisateurs (tant indexeurs qu'internautes) dans leur travail.

CHAPITRE 1

LE DOMAINE COUVERT PAR LE PROTOTYPE

Comme on l'a vu dans l'introduction, le domaine couvert par le thésaurus de l'activité gouvernementale est donc très large tant au niveau des concepts qu'il devra englober que de ses finalités. Le développement d'un tel outil nécessitera des investissements importants et l'implication de plusieurs ressources.

Cependant, et afin de livrer un prototype dans des délais raisonnables, le comité de travail responsable de son élaboration a choisi de réduire le domaine couvert et de concentrer ses efforts sur deux aspects :

- La clientèle : citoyen

Compte tenu des priorités actuelles de rendre accessibles les informations gouvernementales aux citoyens.

- La composante : Pages bleues

Compte tenu que c'est la partie du Répertoire gouvernemental dont la clientèle cible est le citoyen et que ce produit sera disponible sur Internet très prochainement.

Par ailleurs, deux avenues s'offraient pour structurer les termes inclus dans le prototype couvrant les Pages bleues :

- Une structure classificatoire qui tienne compte du vocabulaire découlant de la structure administrative gouvernementale (i.e. des mots significatifs présents dans les noms des unités administratives)
- Une structure théaurale plus thématique et orientée sur les sujets de l'activité gouvernementale.

Compte tenu qu'un outil intégrant les deux approches nécessiterait un investissement trop important, et compte tenu que les noms des unités administratives sont, eux aussi indexés et repérables dans les Pages bleues et qu'ils seront ainsi repérables, il a été retenu que le prototype initial soit plus de nature thématique.



CHAPITRE 2

LE TYPE DE LANGAGE PROPOSÉ POUR LE THÉSAURUS

Compte tenu de la présence de plusieurs indexeurs (autant que de ministères participants), il est apparu essentiel, pour éviter l'éparpillement, que l'indexation soit au moins en vocabulaire contrôlé (par opposition au vocabulaire libre). Cela permettra, entre autres, de garantir une meilleure uniformité et une meilleure cohérence entre les entrées du Répertoire. Ce choix s'explique également par le fait qu'il existera une option de recherche « plein texte ».

Quatre types de langage documentaire à **vocabulaire contrôlé** ont été examinés par le comité, dont deux qui sont pré-coordonnés et deux qui sont post-coordonnés.

2.1 LANGAGES DOCUMENTAIRES À VOCABULAIRE PRÉ-COORDONNÉ

Un vocabulaire pré-coordonné répond au principe suivant lequel l'expression d'un sujet s'effectue au moment de l'élaboration du vocabulaire d'indexation ou au moment même de l'indexation par juxtaposition de termes et création de descripteurs composés, de vedettes-matière, de chaîne de descripteurs, etc. L'expression d'un sujet (avec toutes ses facettes de temps, d'espace, etc.) est la responsabilité de l'indexeur. Les deux types de langages pré-coordonnés examinés sont :

- Les schémas de classification

À titre d'exemple, citons le système Dewey caractérisé par sa classification décimale à structure arborescente :

```

000 Généralités
100 Philosophie
200 Religion
300 Sciences sociales
    370 Éducation
        371 Éducation spécialisée
400 Langage
.....

```

- Les listes de vedettes-matière (comme celle de l'Université Laval).

Il s'agit d'un vocabulaire où l'on retrouve des descripteurs composés formés de plusieurs termes dont la juxtaposition permet d'exprimer à priori un sujet et son contexte par exemple.
Ex. :

```

Femmes – Travail – Québec – 1970
Femmes, Services aux
Arpentage – Canton de Belleville
Langage et histoire

```

2.2 LANGAGES DOCUMENTAIRES À VOCABULAIRE POST-COORDONNÉ

Un vocabulaire post-coordonné répond au principe suivant lequel l'expression d'un sujet s'effectue au moment de la recherche documentaire par combinaison de plusieurs termes ou descripteurs, chacun d'eux ne représentant qu'un seul concept.

Contrairement au vocabulaire pré-coordonné, l'indexeur n'effectue aucune juxtaposition de termes à priori. Chacun des descripteurs représente un concept plus ou moins spécifique, et chacun des descripteurs peut être assigné à un document de façon indépendante. Les deux types de vocabulaires post-coordonnés examinés sont :

- Les thésaurus de descripteurs

Contrairement aux vedettes-matière, les termes d'un thésaurus sont indépendants l'un de l'autre. À titre d'exemple, on pourrait avoir comme termes :

Femmes
 Québec
 Travail
 Arpentage
 Canton de Belleville
 Langage
 Histoire

De plus, les termes du thésaurus s'inscrivent dans une structure hiérarchique (parent/fils) et des relations sémantiques (ex. : synonymes) sont définies entre les termes.

- Les listes de mots-clés

Les termes d'une telle liste sont contrôlés mais ne font l'objet d'aucune relation sémantique (hiérarchie, équivalence, association).

Compte tenu des facilités offertes par les nouvelles technologies pour faire des recherches en associant des termes entre eux (ex. : les opérateurs booléens – et, ou, sauf, ...) et des possibilités de combiner la recherche dans des champs définis et dans les mots du texte intégral, les langages post-coordonnés sont apparus comme étant ceux qui répondaient le mieux aux besoins du Répertoire gouvernemental.

Enfin, les possibilités d'expansion (naviguer du particulier au plus général) et de précision (aller du général au particulier) offertes par le thésaurus ont milité pour cette dernière option plutôt que celle des mots-clés où la navigabilité est plus réduite.

Par ailleurs, et afin d'assurer toute la flexibilité et la convivialité aux indexeurs et aux usagers, il a été prévu que le thésaurus de l'activité gouvernementale sera extensible de manière à pouvoir aménager un espace de travail appelé « mots libres ». Cependant et afin de maintenir la cohérence et l'efficacité du thésaurus et du repérage, les deux principes suivants devront être suivis :

- Principe d'intégrité : un processus de validation devra être implanté pour garantir qu'aucun mot du thésaurus n'est repris par inadvertance parmi les « mots libres ».

- Principe d'évolution : les activités de gestion du thésaurus incluent une analyse fréquente des termes utilisés dans le mode « mots libres » pour voir au respect des règles de bases et chercher à améliorer la cohérence générale tant dans la logique du thésaurus que dans la qualité et l'homogénéité de son utilisation..

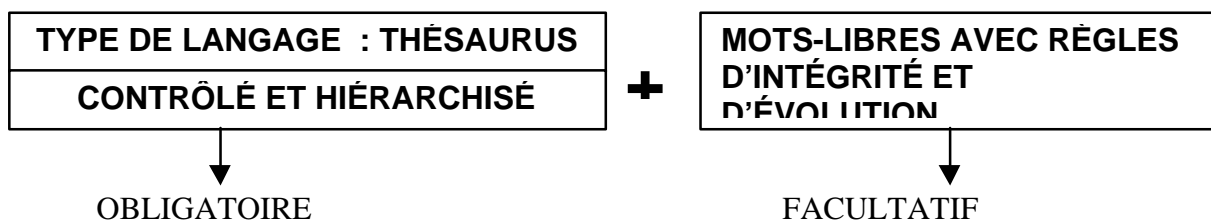
Il sera donc possible d'indexer une fiche (entrée au Répertoire) décrivant un programme ou un service sous deux sections complémentaires :

- **Le thésaurus de l'activité gouvernementale**

Note : Une entrée devra inclure au moins un mot-clé du thésaurus

- **Les mots libres**

Note : Cette section est facultative



CHAPITRE 3

LA STRUCTURE RETENUE POUR LE THÉSAURUS

Le prototype du thésaurus proposé comprend quatre types de relations entre les termes qu'il contient :

- La relation d'équivalence entre termes représentant un même concept (synonymie ; « voir »).
- La relation hiérarchique ou relation d'inclusion entre deux termes (ou lien de paternité et de filiation).
- La relation associative entre différents termes du thésaurus (« voir aussi »).
- La relation de regroupement de termes au sein d'une même catégorie.

3.1 LA RELATION D'ÉQUIVALENCE

Lorsqu'un terme est retenu pour représenter un concept dans le thésaurus, il devient le **terme officiel** utilisable lors de l'indexation. Il peut être intéressant, dès lors, de mettre en relief les synonymes de ce terme, tant pour les indexeurs que pour les utilisateurs.

La relation d'équivalence permet de faire ce type de lien d'un synonyme vers le terme officiel reconnu dans le thésaurus. À titre d'exemple, si on retient le terme « produit agricole » comme terme officiel pour décrire le concept d'un produit résultant de la culture, le terme synonyme « produit végétal » pourrait y être relié par une relation d'équivalence.

Ce type de lien permet, entre autres, le passage du langage naturel au langage artificiel (thésaurus) et permet de réduire le vocabulaire.

Un principe important régit ce type de lien :

- Le synonyme d'un terme officiel représentant un concept dans le thésaurus ne peut pas lui-même être retenu comme terme officiel pour représenter un second concept.

On verra, dans un autre chapitre, que les synonymes sont des clés d'accès possibles lors de la recherche : dans un tel cas, le système indiquera le terme officiel retenu et proposera de faire la recherche avec ce terme.

3.2 LA RELATION HIÉRARCHIQUE

La relation hiérarchique indique un rapport de supériorité ou de subordination entre les termes (ou concepts). Ces liens facilitent la navigation verticale dans le thésaurus. Ils aident les indexeurs et les usagers à atteindre la précision désirée respectivement dans l'indexation et dans la recherche.

Dans l'élaboration du prototype, la relation hiérarchique se limitera aux termes génériques des activités gouvernementales. Les termes spécifiques présents dans les textes (comme, par exemple, des noms de plantes) ont été exclus du thésaurus.

Par ailleurs, s'il advenait qu'un terme très spécifique non inclus dans le thésaurus doit être utilisé pour décrire adéquatement une entrée, il pourra être inscrit dans la section des mots libres.

Trois types de relation hiérarchique ont été retenus pour le thésaurus :

- La relation générique : liant deux termes dans un rapport de genre à espèces. À titre d'exemple, mentionnons « produit laitier et lait ».
- La relation partitive : liant deux termes dans un rapport de tout à partie. Il existe quatre types principaux reconnus pour inclusion dans un thésaurus :
 - Systèmes et organes d'un être vivant.
 - Localisation dans l'espace ou dans le temps.
 - Domaines de la connaissance et disciplines académiques.
 - Hiérarchies sociales, administratives et politiques.

À titre d'exemples, mentionnons « système digestif et estomac », « Amérique du Sud et Venezuela », « Parlement et Sénat ».

- La relation par l'exemple : les termes subordonnés étant des exemples (souvent des noms propres) du terme supérieur. À titre d'exemple, mentionnons la relation entre le terme « pont » et « Pont Pierre Laporte ».

3.3 LA RELATION ASSOCIATIVE

En reliant des concepts entre eux, autrement que par les liens hiérarchiques, ce type de relation enrichit l'indexation et la recherche en associant à un terme d'autres termes pertinents et utiles. Ce type de lien permet la navigation horizontale dans le thésaurus.

À titre d'exemple, on trouve les types de relations suivantes : cause et effet, action et agent, action et produit (ex. : « érable et produit de l'érable »). On peut aussi inclure dans ce type de relations des concepts contraires, s'il devient intéressant de les proposer (ex. : « tolérance et préjugé »).

Contrairement à la règle d'équivalence, les termes associés entre eux doivent obligatoirement être des termes officiels représentant des concepts ; on ne peut donc pas associer à un terme officiel un synonyme ou un terme non officiel.

3.4 LA RELATION DE REGROUPEMENT DE TERMES AU SEIN D'UNE MÊME CATÉGORIE

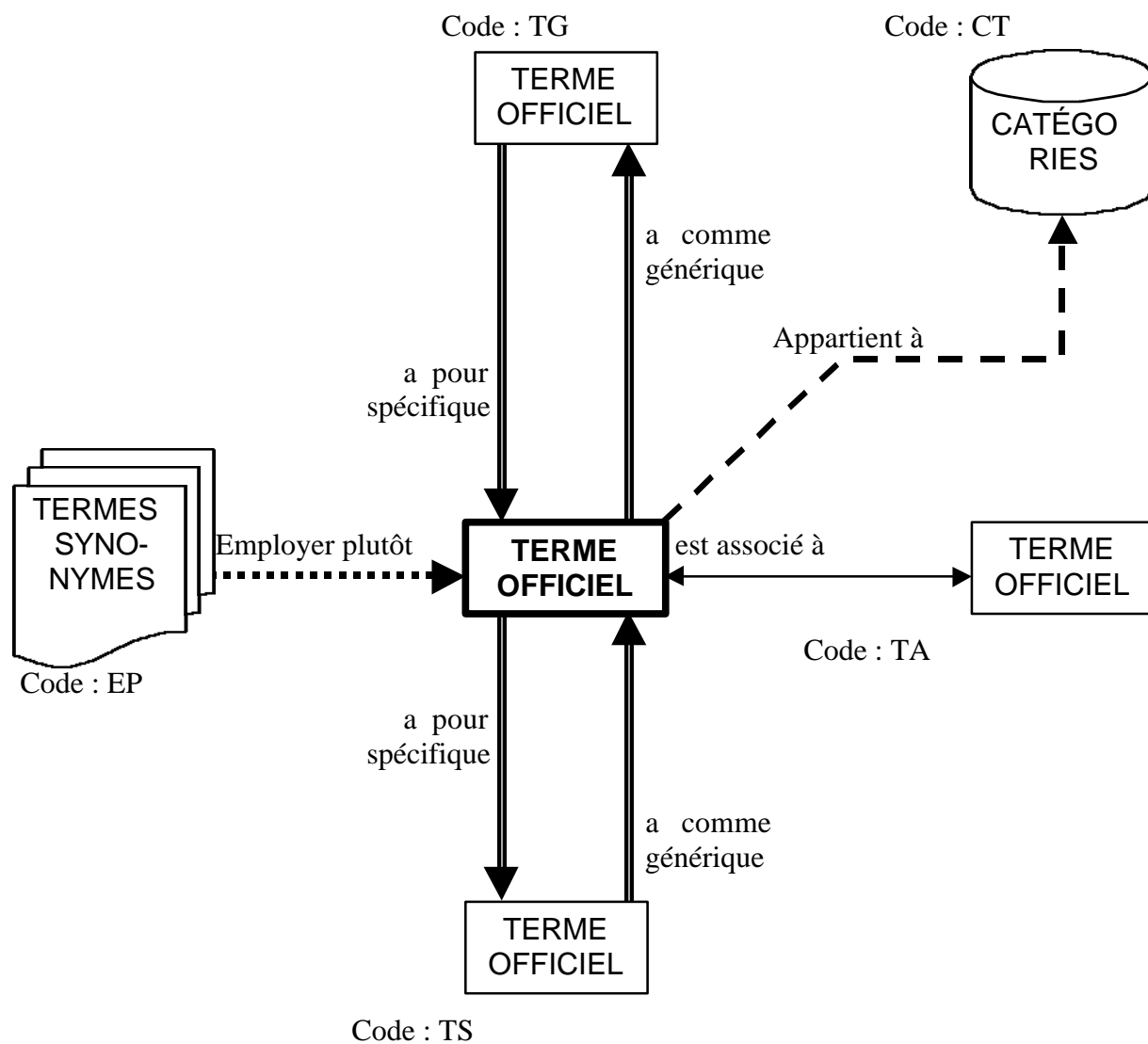
Compte tenu du nombre important prévu de termes supérieurs (embranchements au premier niveau dans la hiérarchie) et afin de faciliter la navigation et le travail des indexeurs, un nombre limité de catégories (la liste est en annexe 1) ont été définies pour regrouper l'ensemble des



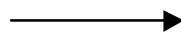

termes officiels du thésaurus. Ces catégories deviennent, en quelque sorte, les unités de regroupement et de présentation des termes officiels du thésaurus. Quatre règles régissent les liens entre les termes officiels et les catégories

- Règle no 1 : Chaque terme accepté dans le thésaurus doit être classé au moins dans une catégorie. Il peut, par ailleurs, être classé dans plus d'une catégorie.
- Règle no 2 : Tout terme dont la relation hiérarchique est de nature "genre à espèce" doit obligatoirement être relié à au moins une des catégories sous lesquelles est classé son terme générique (parent). Il peut, par ailleurs, être classé dans d'autres catégories.
- Règle no 3 : Un terme dont la relation hiérarchique est de nature "tout à partie" ne doit pas obligatoirement être relié à la catégorie sous laquelle est classé son terme générique (parent).
- Règle no 4 : Tout terme dont la relation hiérarchique est de la nature "par l'exemple" doit obligatoirement être relié à la catégorie sous laquelle est classé son terme générique (parent). Il peut aussi être classé dans d'autres catégories.

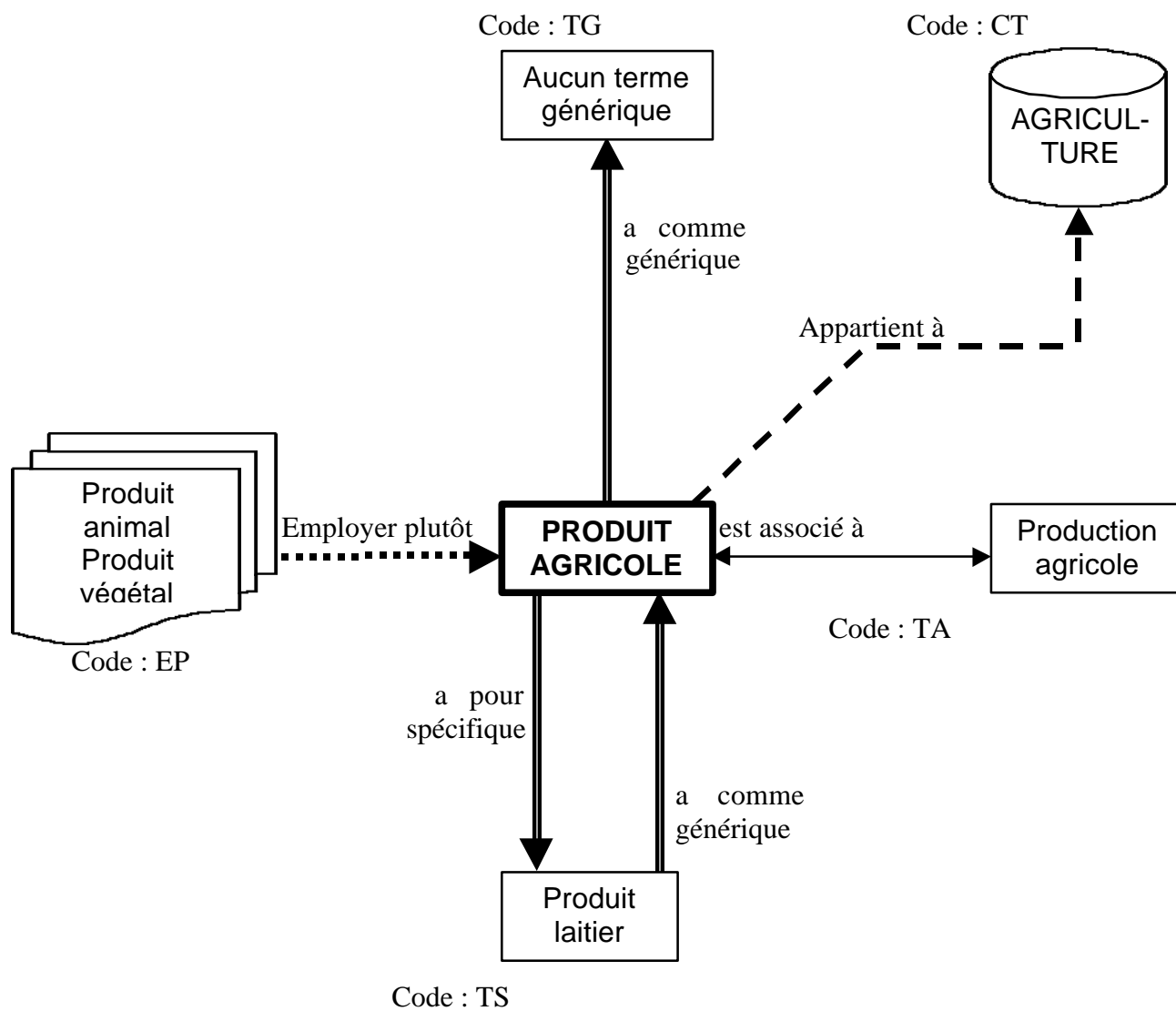
Le schéma de la page suivante représente les liens possibles dans le thésaurus.

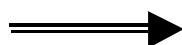



STRUCTURE DU THÉSAURUS



-  Relation hiérarchique (supérieur/subordonné ou père/fils)
-  Relation d'équivalence (synonyme)
-  Relation associative (toujours réciproque entre les deux termes)
-  Relation de regroupement sous une catégorie

La page suivante présente un exemple de ces relations avec le terme « produit agricole ».



-  Relation hiérarchique (supérieur/subordonné ou père/fils)
-  Relation d'équivalence (synonyme)
-  Relation associative (toujours réciproque entre les deux termes)
-  Relation de regroupement sous une catégorie

Dans sa forme manuscrite, on représente les termes de la façon suivante :

PRODUIT AGRICOLE

CT : agriculture
 TS : produit laitier
 EP : produit animal
 produit végétal
 TA : production agricole

Ici, le terme « produit agricole » est un terme officiel du thésaurus. Il n'a aucun terme générique (TG) (c'est à dire qu'il est un nœud supérieur : il n'est pas subordonné à aucun autre terme). Il a, par ailleurs, comme subordonné (TS) le terme « produit laitier ». On indique aussi qu'il est employé pour (EP) « produit animal » et « produit végétal », deux synonymes et on l'associe (TA) au terme « production agricole ». Enfin, ce terme appartient à la catégorie (CT) « agriculture ».

PRODUIT ANIMAL

EM : produit agricole

PRODUIT VÉGÉTAL

EM : produit agricole

Le fait que « *produit animal* » et « *produit végétal* » soient en italique indique qu'il s'agit de termes non officiels du thésaurus. Dans ce cas, il s'agit de deux termes pour lesquels on a préféré le concept « produit agricole ». On doit donc plutôt employer (EM) le terme « produit agricole ». On renvoie donc à ce

PRODUIT LAITIER

CT : agriculture
 TG : produit agricole
 TS : lait
 fromage

Le terme « produit laitier » est un terme officiel du thésaurus qui a comme générique (TG) « produit agricole ». Dans cet exemple, deux termes lui sont subordonnés (TS) : « lait » et « fromage ». Il est sous la catégorie (CT) « agriculture ».

La page suivante reprend le même exemple de façon plus conviviale. Vous noterez que l'on n'y utilise pas les codes (CT, TG, etc.) et que l'on a préféré utiliser directement des expressions plus explicites.

PRODUIT AGRICOLE

Catégorie : agriculture

Spécifique : produit laitier

Employé pour : produit animal
produit végétal

Associé à : production agricole

Ici, le terme « produit agricole » est un terme officiel du thésaurus. Il n'a aucun terme générique (TG) (c'est à dire qu'il est un nœud supérieur : il n'est pas subordonné à aucun autre terme). Il a, par ailleurs, comme subordonné (TS) le terme « produit laitier ». On indique aussi qu'il est employé pour (EP) « produit animal » et « produit végétal », deux synonymes et on l'associe (TA) au terme « production agricole ». Enfin, ce terme est sous la catégorie (CT) « agriculture ».

PRODUIT ANIMAL

Employer plutôt : produit agricole

PRODUIT VÉGÉTAL

Employer plutôt : produit agricole

Le fait que « *produit animal* » et « *produit végétal* » soient en italique indique qu'il s'agit de termes non officiels du thésaurus. Dans ce cas, il s'agit de deux termes pour lesquels on a préféré le concept « produit agricole ». On doit donc plutôt employer (EM) le terme « produit agricole »

PRODUIT LAITIER

Catégorie : agriculture

Générique : produit agricole

Spécifique : lait
fromage

Le terme « produit laitier » est un terme officiel du thésaurus qui a comme générique (TG) « produit agricole ». Dans cet exemple, deux termes lui sont subordonnés (TS) : « lait » et « fromage ». Il est sous la catégorie (CT) « agriculture »

3.5 PARTICULARITÉS

Normalement, il est préférable qu'un même terme officiel ne représente pas deux concepts différents, à moins d'accepter la polyhiérarchie. Pour faciliter l'élaboration du prototype, le comité de travail a décidé d'accepter la polyhiérarchie en autant que le principe suivant soit respecté :

- Lorsqu'on accède à un terme polyhiérarchique, le système devra signaler obligatoirement chaque localisation du terme dans le thésaurus pour permettre à l'utilisateur de choisir un seul concept et ainsi de limiter la recherche au concept qui l'intéresse.

CHAPITRE 4

LES MODALITÉS DE FONCTIONNEMENT À LA DISPOSITION DES INDEXEURS

Quatre modes d'accès seront offerts aux indexeurs :

- Par les catégories du thésaurus

Après avoir choisi une catégorie, on pourra sélectionner un terme apparaissant dans une catégorie et par la suite naviguer à loisir dans le thésaurus, soit dans cette même catégorie grâce aux relations hiérarchiques, soit entre les catégories grâce aux relations associatives.

- Par la liste permutée restreinte des termes du thésaurus. L'accès à la liste se fera en sélectionnant une lettre de l'alphabet. Cette liste inclura les termes officiels ainsi que les synonymes. Après avoir sélectionné un terme, on pourra naviguer à loisir dans le thésaurus.
- Par la liste permutée restreinte des mots libres. L'accès à la liste se fera en sélectionnant une lettre de l'alphabet. On pourra par la suite naviguer dans la liste selon l'ordre de classement des mots libres.
- Par la liste permutée complète combinant les termes du thésaurus et les mots libres. Dépendant du mot sélectionné, l'indexeur se retrouvera dans un des deux modes précédemment décrits.

Dans ces divers modes, l'indexeur n'aura qu'à pointer le terme désiré pour qu'il soit automatiquement inscrit dans l'entrée à indexer. Enfin, il lui sera possible aussi de créer de nouveaux mots libres si nécessaire. Cependant, il ne pourra pas créer de nouveaux termes dans le thésaurus.

CHAPITRE 5

LES MODALITÉS DE FONCTIONNEMENT À LA DISPOSITION DES INTERNAUTES

Lorsqu'un utilisateur fera une recherche par sujet, il aura premièrement, à choisir s'il accède via le thésaurus ou la liste permutée des termes du thésaurus et des mots libres.

- Par le thésaurus

L'utilisateur aura accès, comme pour les indexeurs, soit aux catégories, soit à la liste permutée des termes du thésaurus. Par la suite, il pourra naviguer à loisir dans le thésaurus. Dès qu'il sélectionnera un terme, le système lui indiquera le nombre d'entrées indexées directement sous ce terme ainsi que le nombre d'entrées indexées sous chaque terme (TS) relié au niveau hiérarchique immédiatement inférieur. Un exemple est donné en annexe 2.

- Par la liste permutée des termes du thésaurus et des mots libres.

Dans ce mode, tous les termes du thésaurus sont présents (incluant les synonymes). Cependant, seuls les termes du thésaurus déjà utilisés dans l'indexation des entrées sont inclus dans la liste permutée et ce afin d'éviter de laisser l'utilisateur sélectionner des termes qui ne donneraient aucun résultat.

Dans ce mode, l'utilisateur pourra avoir accès à un menu déroulant (lettres de l'alphabet) ou taper directement le mot qu'il désire. Le résultat sera le suivant :

- S'il a sélectionné un terme du thésaurus, le système lui offrira deux possibilités :
- Soit naviguer dans le thésaurus.
- Soit avoir accès aux entrées indexées, entre autres, sous le terme choisi. Dans ce cas, le système lui indiquera le nombre d'entrées comprenant ce terme ainsi que le nombre d'entrées pour chaque terme spécifique du niveau immédiatement inférieur de la hiérarchie. Un exemple est donné en annexe 2.
- S'il a sélectionné un terme faisant partie des mots libres, le système lui indiquera immédiatement la liste des entrées comprenant ce mot-clé.

CHAPITRE 6

L'ÉLABORATION DU THÉSAURUS

L'élaboration du thésaurus se fait, dans un premier temps, par l'examen des 4 000 descripteurs actuellement utilisés dans la banque de renseignements de Communication-Québec.

La première activité consiste à restreindre le langage aux termes de niveau supérieur, en classant les autres descripteurs pertinents mais plus spécifiques dans une liste qui deviendra, par la suite, la liste initiale des mots libres.

Lorsqu'un terme est retenu pour faire partie du thésaurus, on détermine s'il sera considéré comme terme officiel ou s'il sera considéré comme équivalent (synonyme) d'un terme officiel.

L'étape ultérieure consiste à établir les relations hiérarchiques entre les termes officiels et le regroupement de ceux-ci sous une ou plusieurs catégories.

L'élaboration des relations associatives se fera plus tard lorsque le thésaurus sera assez avancé pour analyser les concepts entre eux et faire ce genre de liens.

Pour supporter ce travail, le logiciel Multites est utilisé. Il s'agit d'un outil spécialisé permettant l'élaboration et la gestion d'un thésaurus. Il gère, entre autres, les quatre types de relations retenues pour le prototype (équivalence, hiérarchique, associative et de regroupement). Il signale aussi les relations conflictuelles entre les termes, supporte la polyhiérarchie et permet l'impression d'un nombre important de rapports de toutes sortes.

Nous prévoyons rendre disponible le prototype pour le mois de septembre 1998.

ANNEXE 1

LISTE DES CATÉGORIES

AGRICULTURE

AMÉNAGEMENT DU TERRITOIRE

CULTURE

ÉCONOMIE

ÉDUCATION

ENVIRONNEMENT

FORESTERIE

GOUVERNEMENT ET VIE POLITIQUE

INDUSTRIE

INFORMATION ET COMMUNICATION

JUSTICE

PÊCHERIES

PERSONNES

RELATIONS INTERNATIONALES

RESSOURCES NATURELLES

SANTÉ

SCIENCE ET TECHNOLOGIE

SOCIÉTÉ

TOURISME ET LOISIRS

TRANSPORT

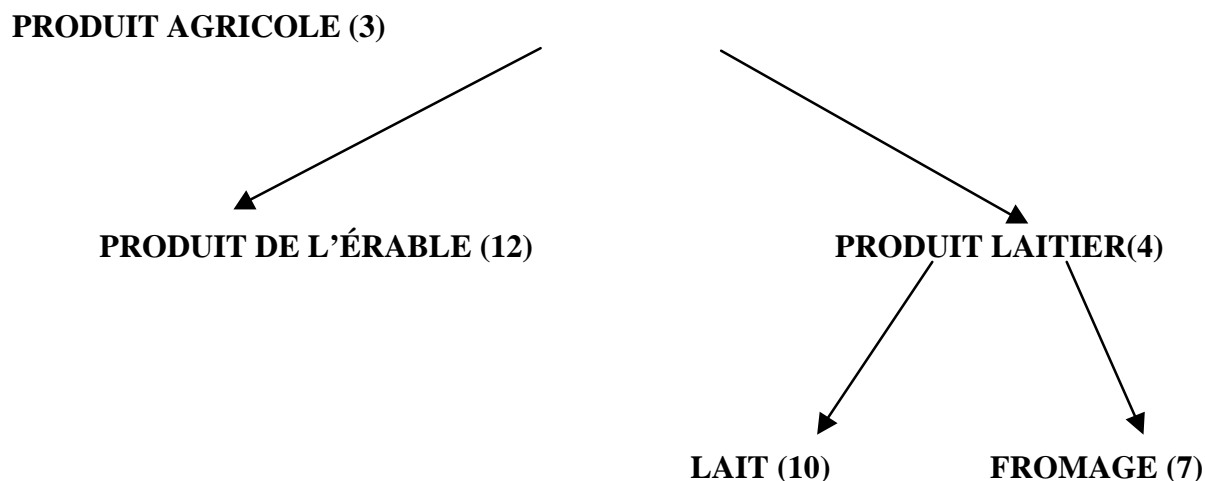
TRAVAIL ET EMPLOI

Cette liste sera revue et ajustée à la lumière des travaux d'élaboration du thésaurus en cours.

Vous noterez que la catégorie « personnes » est une catégorie particulière qui sera détaillée plus tard.

ANNEXE 2

EXEMPLE D'UN RÉSULTAT FOURNI À LA SUITE D'UNE REQUÊTE EFFECTUÉE DANS LE THÉSAURUS



Dans cet exemple, nous comptons 36 entrées en tout sous l'ensemble « produit agricole ». Cependant, 3 de ces entrées seulement traitent des produits agricoles de façon générale. Il y a 12 entrées sous l'ensemble « produit de l'érable » et 21 entrées sous l'ensemble « produit laitier » ; cependant, seulement 4 de ces 21 entrées traitent des produits laitiers de façon générale, 10 traitant du « lait » et 7 du « fromage ».

En sélectionnant, dans le thésaurus, le terme « produit agricole », voici ce qui apparaîtra à l'écran :

Produit agricole, général : 03 entrées
Produit agricole, produit de l'érable : 12 entrées
Produit agricole, produit laitier : 21 entrées