

Un dictionnaire des concepts corporatifs¹

Louis-Claude PAQUIN et Luc DUPUY, chercheurs au
Centre d'Analyse de Textes par Ordinateur
Université du Québec à Montréal

Contexte

Plusieurs contraintes entravent quotidiennement le travail de production et de gestion de l'information textuelle dans les organisations. Le volume et la variété des formats de données textuelles, la complexité de la chaîne documentaire, la diversité des intervenants (et donc des points de vue de lecture), la variété des configurations des plates-formes micro et des logiciels de traitement de texte, sont autant de problèmes liés à la gestion informatique des données documentaires.

S'ajoute à ces problèmes celui du développement terminologique d'une organisation. Il se caractérise par une hétérogénéité et une complexité tributaires de l'accroissement important de la disponibilité de l'information documentaire. Qui plus est, le vocabulaire de l'organisation présente la plupart du temps des divergences et des dénivellations attribuables à l'histoire du développement de l'organisation.

Ces éléments mettent en évidence le rôle non-trivial que doit jouer un cadre intégrateur proprement textuel auquel serait associée une technologie informatique pour l'analyse, la gestion et l'édition des données documentaires. La notion de «poste de travail» qui découlait de programmes dits de quatrième génération, ne recouvre, à notre connaissance, qu'un protocole de communication entre des logiciels génériques : traitement de textes, gestionnaire de bases de données, chiffrier, simulation numérique, etc. Un besoin grandissant se fait de plus en plus sentir pour la mise au point d'un poste de travail normalisé qui intégrerait les différentes fonctions requises spécifiquement pour la gestion des données documentaires (analyse, annotation, tri, classification). Parmi les différentes fonctions que nous jugeons essentielles, il y a la gestion des inventaires lexicaux.

Dans le cadre d'un projet de refonte de la réglementation municipale montréalaise à propos de l'utilisation du sol urbain, nous avons été mandatés pour examiner certains éléments de solution pour effectuer le contrôle socio-terminologique de ce corpus de textes réglementaires. Ce secteur présente les caractéristiques décrites plus haut; les rôles d'intervenants s'échelonnent du préposé à l'information en contact avec le grand public jusqu'à l'architecte. Il nous a semblé opportun en tout premier lieu de constituer un dictionnaire informatisé des concepts corporatifs du client, la Ville de Montréal. Le niveau de notre intervention est celui de la sensibilisation au moyen d'un exemple simple où un seul concept est considéré, celui de «bâtiment». La réalisation informatique est de l'ordre du prototype pour démonstration de faisabilité ce qui représente la phase antérieure à la rédaction du cahier des charges pour une réalisation particulière.

Cette communication présente d'abord la notion de dictionnaire informatisé des concepts corporatifs, puis la méthodologie que nous avons suivie.

¹ À paraître dans les *Actes du colloque sur le repérage de l'information textuelle*, tenu à Montréal le 18 septembre 1991.

Le dictionnaire de concepts corporatifs

Le développement d'un dictionnaire de concepts corporatifs, inspirée de celle de dictionnaire de données (Narayan, R., 1988), est largement utilisée dans le domaine des systèmes de gestion de bases de données. Le dictionnaire de concepts corporatifs nous apparaît apporter une solution efficace et relativement peu coûteuse aux problèmes de l'utilisation du vocabulaire organisationnel. Ce type de développement informatique fait bien ressortir les relations intimes (Meyer, I., Miller, D. & Michaud, D., 1990) existant entre le travail d'analyse terminologique et les méthodologies développées pour systématiser le travail d'acquisition des connaissances bien connues dans l'univers des systèmes experts.

La construction d'un dictionnaire de concepts ne saurait être confondue avec la tâche d'indexation des documents. Cette dernière tâche relève de la bibliothéconomie, et, la plupart du temps, nécessite des moyens considérables et surtout un encadrement par des professionnels de la discipline. L'approche par «dictionnaire de concepts» que nous préconisons offre plutôt une alternative dans les cas où le taux de fréquentation d'un corpus de textes ne justifie pas l'investissement massif de ressources pour l'indexation.

Bien sûr, comme le contenu en est arbitrairement délimité et structuré, sa rigueur dépend de son concepteur. Cependant, lorsque parallèlement à son élaboration, une procédure de validation consistante est mise en place, un dynamique terminologique s'établit au sein de l'organisation. Cette validation peut être effectuée par questionnaires ou entrevues et même par observation. Quels sont les synonymes utilisés par les employés de soutien pour désigner les concepts de l'organisation?

Le dictionnaire de concepts corporatifs ne doit pas être confondu avec un thesaurus et vice versa. D'abord et avant tout, le thesaurus est établi selon les règles de l'art. Quant à lui, le dictionnaire que nous proposons, il ne reposera toujours à la limite que sur des compromis dont certains seront, sans doute, âprement négociés.

Le dictionnaire de concepts corporatifs proposé est implanté avec la technologie des bases de données textuelles, ce qui permet de combiner des formats contraints à des formats plus libres. Cette implantation présente l'avantage qu'en tout temps, l'accès sélectif aux termes et à tous les contextes d'occurrences leur étant associés (concordance) est rapide. Le dictionnaire proposé nous apparaît présenter trois avantages.

Le dictionnaire de concepts corporatifs s'avère un outil efficace pour la planification et la normalisation du vocabulaire corporatif utilisé dans l'organisation. On peut en tirer des extraits pour information ponctuelle, pour rédiger des politiques, des guides et directives, pour expliquer et opérationnaliser la politique terminologique de l'organisation. Ainsi, par exemple, un agent qui rédige un rapport pourra, sans interrompre sa tâche de rédaction accéder à l'ensemble de la terminologie. Le dictionnaire peut devenir un puissant outil de changement socio-terminologique; il permet d'en faire la planification et, par la suite, la validation.

Comme le dictionnaire est fabriqué à partir d'expressions présentes dans un ensemble complet ou représentatif de textes de l'organisation, le respect de la culture corporative est ainsi assuré.

Enfin, l'utilisation d'une technologie proprement textuelle nous assure de la régularité et de l'exhaustivité du processus de construction du dictionnaire. Par ailleurs, la classification des termes repose sur l'utilisateur. Toutefois, la souplesse et la rapidité de rappel des contextes alliées à la convivialité permet aux terminologues ou spécialistes de domaines d'accomplir eux-même et pour eux-même rapidement et de façon économique.

Il est à noter que cette solution est possible dans la mesure de la disponibilité du matériau textuel sur support informatique. Le processus de saisie doit donc être envisagé comme un processus continu.

Un problème important demeure : le passage des savoirs contenus dans les textes, la réglementation municipale dans le cas qui nous occupe, aux structures propres au dictionnaire électronique. Les modalités de ce passage du contenu aux concepts sont cruciales; est-ce une traduction, une conversion ou une adaptation? Ce passage est une opération de nature taxonomique. Après avoir statué sur la délimitation, mais surtout, sur l'admissibilité des termes, ils sont organisés, si possible, sous la forme d'une hiérarchie de la façon la plus stricte possible. En effet, le domaine doit être délimité de façon univoque pour empêcher, ce qui est malgré tout fréquent, l'enchevêtrement des hiérarchies relevant d'univers conceptuels différents.

Il s'agit là certainement d'un des principaux obstacles à l'informatisation des savoirs (Boose, J & Gaines, B, 1990; Berry, & Hart, 1990; Paquin 1991). Ce type d'obstacle présente deux ordres consécutifs de difficultés :

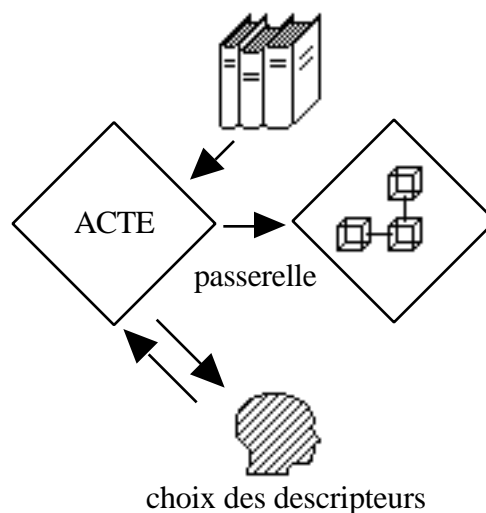
D'abord, la chaîne de traitements pour représenter les savoirs dans un formalisme simple et rigoureux.

Ensuite, le processus de diffusion, de consommation et de mise-à-jour de la base de connaissance.

Bien qu'il reste passablement de chemin à parcourir sur la voie de l'indexation automatique, un dispositif computationnel est quand même envisageable immédiatement (Genuardi, M. 1990) à condition de laisser à l'humain le choix des descripteurs. C'est ce que nous proposons.

Le dispositif proposé

Pour les besoins de la présente démonstration, le premier ordre de difficulté est contourné, car le dispositif informatique proposé est générique, c'est-à-dire qu'il n'incorpore aucun savoir ni présupposé quant au domaine de référence. Si l'aspect gestion des données de la tâche est entièrement prise en charge, les aspects évaluatif et taxonomique sont laissés à l'utilisateur. Le dispositif que nous avons mis au point s'applique donc aux phases antérieure et postérieure à la prise de décision comme le montre le schéma suivant :



La phase antérieure en est une d'analyse textuelle alors que la phase postérieure consiste en la projection des descripteurs sélectionnés sur le texte et l'organisation hypertextuelle des segments dépistés qui constituent les entrées du dictionnaire. Quant au second ordre de difficulté, il est minimisé par trois facteurs:

- la souplesse de la technologie des bases de données textuelles;
- une stratégie soutenue (gestion des mises-à-jour) de diffusion;
- la pertinence effective du service rendu par le dispositif à l'organisation.

En contre-partie de l'automatisme, nous offrons une fonction de lecture permanente qui assure à l'utilisateur, à toutes les fois qu'il aura besoin de prendre des décisions, l'exhaustivité des contextes d'occurrences et de leurs annotations. La technologie proposée est proprement textuelle. Elle permet d'incorporer une quantité maximale de dimensions : onomastique, référentielle, thématique, etc. Un usage récursif des éléments du dictionnaire de concepts corporatifs permettrait, en effet, d'obtenir une stabilité de la terminologie par un processus continue réglé de d'adaptation des termes aux réalités du domaine : une politique terminologique en quelque sorte.

Avant de décrire chacune des phases du dispositif, nous présentons l'environnement informatique utilisé. L'Atelier Cognitif et Textuel (ACTE, Daoust, F., Paquin, L.-C., Dupuy, L., 1989) est dédié aux problèmes d'analyse de données textuelles et d'élaboration de systèmes d'information. Il comporte deux modules principaux : a) un système d'analyse textuelle et b) un générateur de systèmes à base de connaissances auxquels s'ajoutent une série d'utilitaires de pré et post-traitements de textes développés à l'aide des langages AWKTM (Aho, A., Kernighan, B et Weinberger, P., 1988) et ICONTM (version 8, Griswold, R., 1990).

Le dépistage des concepts ou descripteurs qui constitueront le dictionnaire repose en bonne partie sur la possibilité d'analyser en détail la terminologie d'un corpus de textes représentatifs du domaine. Pour ce faire, l'ACTE offre un Système d'Analyse de Textes par Ordinateur (SATO version 3.6, Daoust, F. 1990) pour manipuler les aspects lexicaux et syntagmatiques des textes. On peut ainsi indexer automatiquement les éléments de la morphologie grammaticale (noms, adjectifs, prépositions, etc.) à l'aide des Bases de Données Lexicales (Dupuy, L, 1989).

Une grammaire des segments nominaux permet de faire des extractions du type Nom + préposition + Nom, Adjectif + Nom, etc., (Bouché, R, 1989). L'analyse des résultats doit être faite par un expert de domaine afin de dégager les lexies complexes pertinentes. Cette consolidation terminologique effectuée, il reste à les classer. Assez paradoxalement, même si la classification est une des opérations fondamentales de la cognition, on sait finalement peu de choses sur le sujet. Ici l'empirie est la seule méthode. On observe et on propose des regroupements potentiels qui sont par la suite renforcés ou réfutés.

La classification est en fait l'étape ultime d'une processus qui comporte auparavant des opérations de décontextualisation, d'explicitation, de condensation et surtout de de standardisation des segments. Le processus de classification est conçu ici comme un mécanisme complexe d'annotations textuelles, annotations déposées par des dictionnaires et des algorithmes d'analyse de textes pour modéliser les différentes structures du texte : lexicales et segmentales. Contrairement aux conséquences d'une méthodologie conventionnelle qui aurait pour effet en quelque sorte de figer l'ensemble des questions que l'on peut poser à un corpus de textes. Notre approche permet en tout temps de modifier les descripteurs et d'obtenir rapidement un étiquetage qui tienne compte des changements apportés au dictionnaire.

La suite des opérations

La première opération consiste à opérer une sélection préliminaire sur la base des concepts principaux du domaine. L'intuition doit préférablement provenir d'un expert du domaine. Dans notre cas, à titre d'illustration, le concept de bâtiment a été sélectionné. Un contexte délimité par les phrases a été demandé. Une correspondance de la phrase avec les alinéas a été constatée dans la réglementation municipale que nous avons traité :

- Nonobstant la stipulation du premier paragraphe de l'article 1, il est permis d'occuper ou d'utiliser tout lot ou terrain, et de construire, réparer, modifier, transformer, occuper ou utiliser tout bâtiment: 1° pour fins d'éducation et d'enseignement dans le territoire borné par la rue Sherbrooke, la ligne d'arrière des lots aboutissant au côté est de la rue University, la limite du parc Mont-royal et la rue Peel; 2° pour fins d'hôpitaux et de dispensaires dans la partie de ce territoire située au nord de l'avenue Pine.

2 *page=1651-01b

- Nonobstant la stipulation du premier paragraphe de l'article 1 ci-dessus, il est permis au Bureau International du Travail d'occuper pour ses propres fins le bâtiment portant le no 3450 de la rue Drummond et ses dépendances, situés sur le lot no 1758-5 du cadastre du quartier Saint-antoine, pourvu que ces bâtiments ne soient en aucune façon agrandis et qu'il ne soit construit aucun bâtiment additionnel sur ce lot.

La deuxième opération consiste en l'annotation des concordances. Lors d'un balayage, on doit repérer les expressions qui sont conceptuellement associées au concept qui fait l'objet d'une description, ici «bâtiment» ci-dessous représenté en relief. Les expressions repérées sont mis en caractère gras :

1 *page=1651-01a

- Nonobstant la stipulation du premier paragraphe de l'article 1, il est permis d'occuper ou d'utiliser tout lot ou terrain, et de construire, réparer, modifier, transformer, occuper ou utiliser tout **bâtiment**: 1° pour fins d'éducation et d'enseignement dans le territoire borné par la rue Sherbrooke, la ligne d'arrière des lots aboutissant au côté est de la rue University, la limite du parc Mont-royal et la rue Peel; 2° pour fins d'hôpitaux et de dispensaires dans la partie de ce territoire située au nord de l'avenue Pine.

2 *page=1651-01b

- Nonobstant la stipulation du premier paragraphe de l'article 1 ci-dessus, il est permis au Bureau International du Travail d'occuper pour ses propres fins le **bâtiment** portant le no 3450 de la rue Drummond et ses dépendances, situés sur le lot no 1758-5 du cadastre du quartier Saint-antoine, pourvu que ces **bâtiments** ne soient en aucune façon agrandis et qu'il ne soit construit aucun bâtiment additionnel sur ce lot.

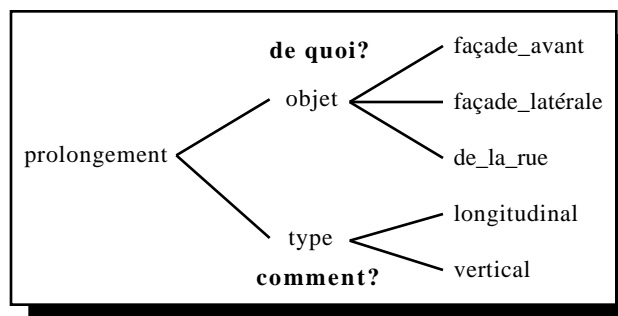
Le rythme du balayage peut être soutenu en raison de la saturation lexicale du corpus indiquée par une grande redondance des mots. La modalité d'association est duelle : les expressions syntaxiquement associées au concept étudié, et celles qui sont dans le voisinage. Voici une liste d'expressions liées syntaxiquement au concept de «bâtiment»; celles qui pourraient s'avérer significatives ont été imprimées en gras :

1	bâtiment <u>principal</u> situé
27	bâtiment <u>projeté</u>
5	bâtiment <u>projeté</u> de façon
1	bâtiment <u>reconstruit</u>
11	bâtiments <u>à caractère temporaire</u>
1	bâtiments affectés
1	bâtiments à ligne latérale
1	bâtiments <u>augmentés en hauteur</u>
1	bâtiments <u>autorisés</u>
1	bâtiments <u>commerciaux</u>
1	bâtiments <u>d'enseignement</u>
2	bâtiments <u>en groupe continu</u>
1	bâtiments <u>pour fins de culte</u>
1	bâtiments <u>recyclés</u>
1	bâtiments <u>sous forme de tours</u>

Le tableau suivant illustre l'utilisation des fréquences de la co-occurrence comme indice de voisinage. On remarque dans l'exemple suivant que le concept de prolongement est fortement associé au concept de bâtiment. Ce concept est de niveau intermédiaire parce qu'il est lui-même l'objet d'une structuration par les expressions en gras :

freq cooc.		
21	11	prolongement
2	2	prolongement <u>de la façade avant</u>
1	1	prolongement <u>de la façade latérale</u>
2	2	prolongement <u>des façades avant</u>
1	1	prolongement <u>longitudinal</u>
7	4	prolongements
1	1	prolongements <u>de la rue</u>
11	4	prolongement <u>vertical</u>

À titre d'illustration, voici une ébauche de structuration du concept de «prolongement» qui gravite autour de celui de «bâtiment». Les catégories qui forment le niveau intermédiaire, ici en gras, sont inférées.



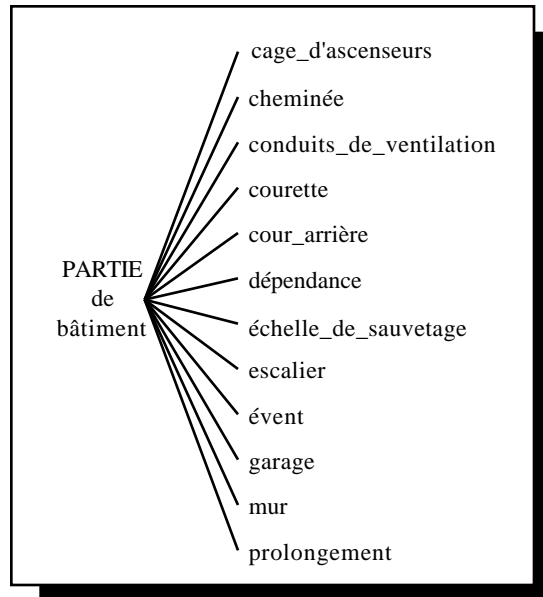
Il vaut mieux, à notre avis, opter pour des catégories simples basées sur une justification extérieure à la problématique. À titre d'illustration, nous avons regroupé un certain nombre de verbes d'actions; la catégorie est nommée «ACTION». Plus simples et évidents pour tous sont les critères, plus robuste est la taxonomie.

occuper
utiliser
lot
terrain
construire
réparer
modifier
transformer
occuper
utiliser
fins d'éducation et d'enseignement
ligne d'arrière des lots
fins d'hôpitaux et de dispensaires

À titre d'illustration supplémentaire d'une taxonomie basée sur une analyse terminologique, l'exemple suivant regroupe sous une même catégorie, appelée «caractéristiques du bâtiment».

alignement_frontal
front_sur_une_voie_publicque_d'_emprise_supérieure
dégagements_latéraux
éclairage
éloignement_entre_volumes_distincts
espaces_réservés_aux_garages
recul_arrière
implantation_maximale
nombre_d'étages_permis
nombre_maximum_de_logements
nombre_maximal_de_places_de_stationnement_autorisées_à_l'extérieur
occupation_projetée
projection_horizontale_des_parties
recul_par_rapport_à_la_voie_publicque
superficie_totale_du_plancher

Il est à noter que dans les textes d'où sont extraites les expressions regroupées, une liaison syntaxique, habituellement de type détermination, les unit à une valeur possible. Ainsi, l'ensemble des déterminants liés syntaxiquement avec chacun des concepts constitue en quelque sorte un domaine de valeur pour ce concept. À titre d'illustration nous avons retenu la caractéristique «espaces réservés aux garages». Le garage en tant que tel nous intéresse parce qu'il apparaît sur la liste des cinquante parties de bâtiment qui sont mentionnées au travers l'ensemble des textes.



Une revue des trente contextes d'occurrence du mot «garage» s'avère très instructive sur ce concept. On apprend notamment la définition, la norme de la marge latérale et ses dérogations, la norme d'utilisation et ses dérogations :

3470-02-01
Garage particulier: toute dépendance isolée ou contiguë au bâtiment principal ou toute partie de celui-ci servant au remisage des automobiles.

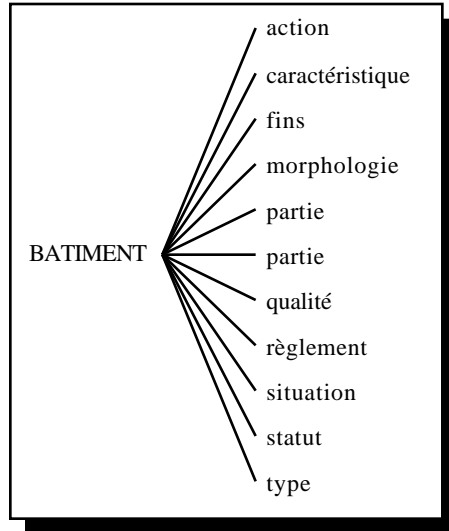
3722-03.4
Un garage particulier ou un abri-auto isolé doit être distant d'au moins six (6) pieds et six (6) pouces du bâtiment dont il constitue une dépendance.

3722-03.4
(6924 art.2, 24-02-1986) [Lorsqu'un garage particulier d'un étage ou un abri-auto est contigu à une maison d'habitation d'au plus trois (3) logements, la **marge latérale** le long du garage ou de l'abri-auto peut être réduite à quatre (4) pieds (1,2 m).

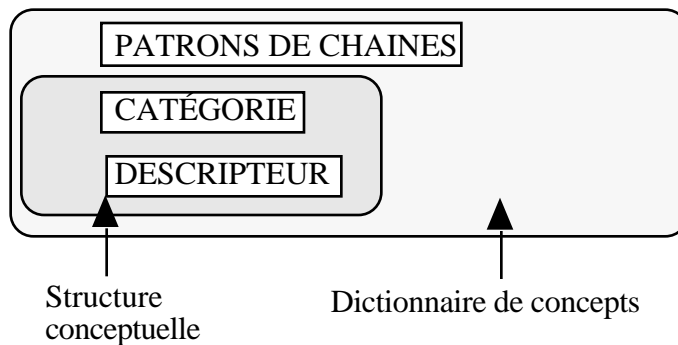
3810-06-07
- Stationnement
Dans la zone d'habitation, il est interdit d'utiliser tout garage particulier, abri-auto, terrain vacant ou espace libre sur un terrain déjà bâti pour le stationnement des véhicules automobiles autres que les véhicules automobiles de promenade.

3810-06-07
(5591 art.15c, 12-12-1980) []
En exception aux dispositions du présent article, le propriétaire d'un auto-taxi, tel que défini à l'article 1 du règlement 2745, peut garer ce véhicule dans un garage particulier ou un abri-auto de la maison qu'il habite.

Lorsque l'ensemble des expressions retenues ont été catégorisées de façon satisfaisante, la taxonomie se stabilise. Voici, avant approbation par les experts du domaine, les classes qui ont été empiriquement retenues pour décrire le concept de «bâtiment» :



La méthodologie empirique, à partir de l'étude des contextes d'occurrences, que nous proposons reconstruit des structures conceptuelles. Au lieu de ne retenir du processus de structuration que la structure elle-même, le dictionnaire de concepts énumère pour chacune des catégories induites les réalisations lexicales pertinentes, ou tout au moins les prédit à l'aide d'un patron de chaîne de caractères, comme le montre le schéma suivant :



Voici, à titre d'illustration, une liste bipartite où la catégorie est accolée aux réalisations lexicales de cette catégorie.

CAT EXPRESSION

a02 type maison_d'appartement
b02 qualité additionnel
b03 qualité principal
c01 statut d'_origine
c02 statut démoli
c03 statut de_caractère_temporaire
c04 statut projeté
c05 statut vacant
c06 statut recyclé
c07 statut annexé
d01 situation à_ligne_latérale_zéro
d02 situation adjacent
d03 situation adossé_d'un_côté

Par la suite, les descripteurs retenus sont transformés en filtres qui seront projetés sur le corpus des textes, de sorte que tous les segments pertinents seront étiquetés. Procéder par patrons pour effectuer la catégorisation procure une garantie d'exhaustivité dans la mesure de la couverture du patron formulé. Ici plusieurs stratégies sont envisageable, telles la troncature pour déjouer les variations morphologiques, la recherche dans un contexte élargi pour dépister les emboîtements potentiels, etc. La mise en application de chacune de ces stratégies entraîne un taux de silence (information pertinente non-dépistée et un taux de bruit (information dépistée qui s'avère inutile) qui dépend du type de textes. L'état suivant de la liste précédente montre en caractères gras une troncature morphologique, «démoli\$», marquée par le signe «\$»; les expressions dépistées peuvent autant être masculin que féminin, singulier que pluriel. Il montre aussi une troncature sémantique, «proj\$»; la gamme des expressions dépistées est alors élargie à {«projeté - e -s», «à l'état de projet», «projection»}. Si les deux premières expressions peuvent être catégorisées «statut», il n'en va pas de même pour le troisième qui relèverait plutôt d'une catégorie comme «vue»

abréviation 0 concordance stricte
abréviation 1 *desc:

*0maison_d'appartement*1a02
*0additionnel*1b02
*0principal*1b03
*0d'_origine*1c01
*0**démoli**\$*1c02
*0de_caractère_temporaire*1c03
*0**proj**\$*1c04
*0vacant*1c05
*0recyclé*1c06
*0annexé*1c07
*0à_ligne_latérale_zéro*1d01
*0adjacent*1d02
*0adossé_d'un_côté*1d03

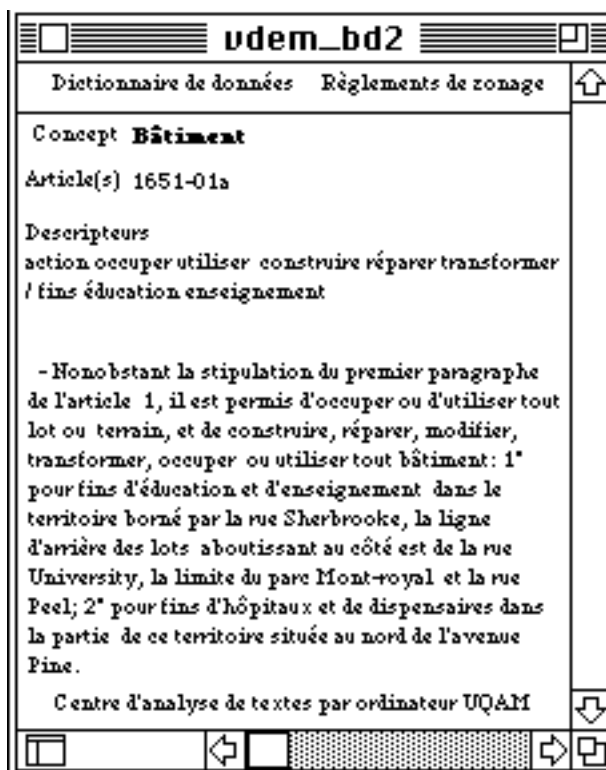
La seconde étape de notre dispositif consiste à rendre disponible dans une base de données textuelle les résultats de l'annotation du corpus. Pour ce faire, les patrons sont projetés sur le corpus d'origine, un ensemble de règlements de zonage, ou encore, ils sont projetés sur des corpus connexes, comme les règlements généraux. Nous reprenons ici le premier contexte d'occurrences du début :

1651-01a
 - Nonobstant la stipulation du premier paragraphe de l'article 1, il est permis d'occuper***desc=i08** ou d'utiliser***desc=i12** tout lot ou terrain, et de construire***desc=i02**, réparer***desc=i10**, modifier***desc=i11**, transformer***desc=i11**, occuper***desc=i08** ou utiliser***desc=i12** tout bâtiment:
 1° pour fins d'éducation***desc=j07** et d'enseignement***desc=j06** dans le territoire borné par la rue Sherbrooke, la ligne d'arrière des lots aboutissant au côté est de la rue University, la limite du parc Mont-royal et la rue Peel;
 2° pour fins d'hôpitaux et de dispensaires***desc=j07** dans la partie de ce territoire située au nord de l'avenue Pine.

Les passerelles qui ont été développées pour les fins de la présente démonstration de faisabilité appartiennent au domaine public. Nous les avons de plus documentées. Le traitement consiste à retirer du texte les marques de descripteurs, à copier la réalisation lexicale, puis écrire les catégories avec leur valeur dans le contexte :

1651-01a
 - Nonobstant la stipulation du premier paragraphe de l'article 1, il est permis d'occuper ou d'utiliser tout lot ou terrain, et de construire, réparer, modifier, transformer, occuper ou utiliser tout bâtiment: 1° pour fins d'éducation et d'enseignement dans le territoire borné par la rue Sherbrooke, la ligne d'arrière des lots aboutissant au côté est de la rue University, la limite du parc Mont-royal et la rue Peel; 2° pour fins d'hôpitaux et de dispensaires dans la partie de ce territoire située au nord de l'avenue Pine.
action occuper utiliser construire réparer transformer
fins éducation enseignement hôpitaux et dispensaires

Voici la représentation du précédent fichier après importation. Chacun des champs est indexé et peut donc être fouillé efficacement. L'inconvénient c'est que les tables d'adresses lexicales, utilisées par le programme pour accélérer la fouille, occupent plus de place que le texte lui-même :



L'utilisateur peut, à volonté, interroger soit les éléments lexicaux du corpus et leur réalisation au niveau des segments physiques du texte, soit les points de vue conceptuels associés aux segments, soit des contextes définitoires, explicatifs, illustratifs, etc. Il reste cependant que ce type de bases de données textuelles n'offre pas toute la souplesse qu'exige la lecture d'un dictionnaire de concepts corporatifs.

L'édition hypertextuelle représente, à notre avis, une technologie plus appropriée à l'exploitation des résultats. Un hypertexte ou hyperdocument (Balpe, J.-P., 1990) est, rappelons-le, un contenu informatisé offrant la possibilité d'être informé à partir de multiples perspectives de lecture. La lecture hypertextuelle est une stratégie qui permet non seulement d'examiner le comportement d'un élément particulier du dictionnaire dans le corpus à partir de paramètres transversaux, mais aussi d'y déposer ses propres annotations : descripteurs, commentaires, liens avec des textes hors corpus, etc. Ainsi, toutes les traces laissées par les lectures expertes peuvent être réutilisées pour accéder aux items appropriés.

Conclusion

À faible coût d'investissement initial, le dictionnaire des concepts corporatifs génère deux types de retombées : d'abord il matérialise la structure organisationnelle de la terminologie, ici du domaine municipal. Puis, le dictionnaire devient, en quelque sorte, une interface permettant à l'utilisateur de naviguer dans le corpus de textes. Il va sans dire que ce processus n'épuise pas la complexité des opérations de la chaîne documentaire, mais il peut par soi-même à faible coût, dans la mesure de la disponibilité des textes, rendre des services appréciables et parfois même suffisants.

En terminant, le dictionnaire de concept est un élément sur lequel on peut s'appuyer pour garantir la rigueur et l'exhaustivité du développement de systèmes à base de connaissances dans le secteur des institutions publiques et para-publiques.

Bibliographie

Aho, A., Kernighan, B et Weinberger, P., *The AWK Programming Language*, Addison-Wesley Publishing Company, 1988.

Balpe, J.-P., *HYPERDOCUMENTS, HYPERTEXTES, HYPERMEDIAS*, Eyrolles, Paris, 1990.

Berry, & Hart, *Expert systems, Human issues*, The MIT Press, Cambridge, Massachussets, 1990.

Boose, J & Gaines, B, *The Foundations of Knowledge Acquisition*, Academic Press, Harcourt Brace Jovanovich, Publishers, New York, 1990.

Bouché, R., "Le syntagme nominal, une nouvelle approche des bases de données textuelles", *META*, Vol. 34, N° 3, Septembre 1989, Les Presses de l'Université de Montréal, pp. 428-434.

Daoust, F., Paquin, L.-C., Dupuy, L., ACTE : "Workbench for Knowledge Engineering and Textual Data Analysis in the Social Sciences", *Proceedings fo the Fourth International Conference on Symbolic and Logical Computing*, Dakota State University, Madison, South Dakota, 1989.

Dupuy, L., *Les Bases de données lexicales*, Document de travail, Centre d'Analyse de Textes par Ordinateur, Université du Québec à Montréal, 1991.

Genuardi, M., "Knowledge-based Machine Indexing from Natural Language Text: Knowledge Base Design, Development and Maintenance", *Proceedings of the Second International Congress on Terminology and Knowledge Engineering Applications*, Indeks Verlag, Frankfurt, 1990, pp. 345-351

Griswold, R. & Griswold, M., *The Icon Programming Language*, Prentice Hall, Englewoods Cliffs, New Jersey, 1990.

Narayan, R., *Data Dictionary, Implementation, Use, and Maintenance*, Prentice Hall, Mainframe Software Series, Englewoods Cliffs, New Jersey, 1988

Paquin, Louis-Claude, "Du terme au concept", *Actes du Colloque international 'Les industries de la langue: Perspectives des années 1990'*., Office de la langue française, Société des traducteurs du Québec, Montréal pp. 313-333.