Le travail conceptuel collectif : une analyse assistée par ordinateur du concept d'ACCOMMODEMENT RAISONNABLE dans les journaux québécois

Jean-François Chartier, Jean-Guy Meunier, Jean Danis et Mohamed Jendoubi

UQÀM - LANCI (Laboratoire d'Analyse Cognitive de l'Information) C.P. 8 888
Succ. Centre-Ville – Montréal – Québec – Canada, H3C 3P8

Abstract

Computer assisted conceptual analyisis of texts (CACAT), is a method to assist the interpretative analysis of a concept in a textual corpus, produced by an individual or a community. We present at first the hypotheses and the methodological steps establishing (constituting) a CACAT chain. In the second step, we present research results on the social distribution of the conceptual work of "ACCOMMODEMMENT RAISONNABLE" in the Quebecois newspapers. This analysis aims at discovering the set of semantic and inferential properties associated with this concept.

Résumé

La lecture et l'analyse conceptuelle de textes assistée par ordinateur (LACTAO), est une méthode pour assister l'analyse interprétative des concepts dans un corpus de textes, produit par un individu ou une communauté. Nous présentons dans un premier temps les hypothèses et les étapes méthodologiques constituant la chaîne LACTAO. Dans un deuxième temps, nous présentons quelques résultats préliminaires de recherche sur la distribution sociale du travail conceptuel sur les ACCOMMODEMENTS RAISONNABLES dans les journaux québécois. Cette analyse montre comment se déploie, sur différents vecteurs de sens, le concept étudié.

Mots-clés : LACTAO, interprétation, concept, concordance, classification, annotation, ACCOMMODEMENT RAISONNABLE.

1. Introduction : l'interprétation conceptuelle des textes

Pour les chercheurs des humanités et des sciences humaines et sociales, l'analyse et l'interprétation conceptuelle des textes constituent une étape importante de leur démarche scientifique. Philosophes, historiens, sociologues, anthropologues, sont tous amenés, à un moment ou un autre, à une étape d'interprétation conceptuelle des textes : pour certains des textes philosophiques, pour d'autres de textes religieux, historiques, des articles de journaux, des transcriptions d'entrevues, etc. (Goody 1979).

Ce type d'analyse dans les diverses disciplines des sciences humaines et des lettres trouve ses fondements et ses outils dans plusieurs sources : linguistiques, théories de l'énonciation et du discours, analyse qualitative, philologie classique et bien d'autres.[1] Pour notre part, dans cet article, nous définirons l'analyse conceptuelle comme une méthode interprétative pour

[1] L'analyse conceptuelle des textes prends plusieurs formes selon les disciplines. Si le philosophe parle généralement du concept de concept (Meunier 2006), l'anthropologue parle davantage de « schéma culturel » (D'Andrade 1995), le psychosociologue de « représentation sociale » (Jodelet 1989), certains d'idéologie, de stéréotype, etc.

l'exploration systématique des propriétés sémantiques et du réseau d'inférences d'un ensemble de prédicats exprimant un concept particulier dans un texte ou un discours (Meunier 2006 ; Brandom 1994 ; Rey 1983).

Dans le domaine des sciences humaines et les lettres, on trouve certains outils informatiques d'assistance à cette analyse interprétative experte : les systèmes d'analyse qualitative de contenu des corpus de textes, tels Atlas et Nudist (Barry 1998 ; Wilson 2001), les outils d'analyse de textes d'inspirations statistiques (Lebart et Salem 1997) tel Alceste (Reinert 1994). Du côté étasunien et canadien, on trouve des outils du type « forage de texte » (Unsworth 2005 ; Hearts 1998 ; Meunier et al. 2005 ; Alexa et Zuell 1999), mais avec des stratégies mathématiques légèrement différentes.

Ces outils assistent l'identification de relation, tantôt dite thématique, tantôt sémantique, sur l'ensemble des mots (simples ou composés) de l'ensemble d'un corpus. Bien qu'heuristiques, utilisées tel quelles, ces stratégies d'analyse permettent difficilement au lecteur-analyste expert de diriger l'attention sur un concept pôle et d'en explorer finement le fonctionnement.

Nous proposons ici une méthodologie sous forme de chaîne de traitement informatique, qui s'inspire du même horizon mathématique et linguistique, mais dont le but est d'assister par ordinateur l'analyse et l'interprétation d'un concept précis dans un corpus de textes. Cette méthode peut être appliquée à l'analyse et l'interprétation d'un concept chez un auteur particulier ou appliqué à un concept socialement partagé au sein d'une communauté épistémique.

Nous présentons une application concrète de cette méthode, appelée la Lecture et l'Analyse Conceptuelle de Textes Assistée par Ordinateur (LACTAO). Les hypothèses de la méthode seront d'abord présentées. Ceci est suivi de quelques résultats préliminaires d'une interprétation conceptuelle d'inspiration sociologique, sur le travail cognitif collectif de conceptualisation à propos des ACCOMMODEMENTS RAISONNABLES, dans les journaux québécois.[2]

2. LACTAO : Une méthode assistée par ordinateur pour l'interprétation conceptuelle experte des textes dans les sciences humaines et sociales

LACTAO est une stratégie informatique, qui permet d'étudier via une chaîne de traitement constituée de classifieurs mathématiques (neuronaux), la distribution d'une expression linguistique spécifique dans un corpus de textes. Elle s'adresse aux lecteurs-analystes *experts* des humanités, des sciences sociales et humaines, qui font de l'interprétation conceptuelle des textes, un moment important de leur démarche scientifique.

Plusieurs stratégies peuvent être utilisées pour la réaliser. Récemment, certains (Loiseau 2005 ; Vallette, 2003, Rastier 2005,) ont exploré des approches sémantico-sémiques. Mais ces types d'analyses sont surtout centrés sur les dimensions lexicales et thématiques des textes. L'approche est donc surtout de nature linguistique. Or, dans sa pratique classique, l'analyse conceptuelle explore certes la dimension lexicale, son contenu sémantique et ses conditions d'énonciation, mais elle vise plus, à savoir : le réseau des propositions qui participent à la

[2] Nous avons ailleurs appliqué cette même méthode sur des concepts de type philosophique — le concept de MIND chez Pierce (Meunier et Forest 2008); le concept de ÉVOLUTION chez Bergson (Danis et Meunier 2008).

constitution du sens de ces prédicats (sa variation intensionnelle) et des inférences que l'on peut en tirer (variation inférentielles).[3]

Un tel type d'analyse ne peut se contenter d'utiliser des approches trop générales ou trop descendantes (top down), qu'offrent par exemple les outils de type moteur de recherche, les thésaurus, l'analyse thématique, le forage de texte, certaines analyses distributionnelles ou statistiques. Plus de finesse et de précision sont requises. Ce fut un des reproches majeurs que la communauté des lettres a apporté à l'analyse de textes assistée par ordinateur (Rockwell 2003) : elles n'assistent pas l'analyse experte requise. [4]

LACTAO vise à mieux se conformer aux exigences de cette interprétation experte du concept. Cela se traduit en deux cheminements distincts : premièrement, elle assiste l'analyse *micro* des propositions qui définissent le contenu conceptuel d'une *expression spécifique*, chez un auteur particulier ou distribué chez une communauté épistémique ; deuxièmement, elle offre une assistance de type ascendante (bottom-up), qui d'une part ne repose pas sur une connaissance a priori des propriétés du ou des concepts étudiés, et d'autre part, permet la découverte de ces propriétés.

3. Hypothèse générale

LACTAO se présente donc comme une assistance informatique à l'analyse conceptuelle des textes, qui veut respecter les exigences liées au travail interprétatif *expert.* Ceci en s'appuyant sur l'hypothèse générale suivante : *L'expression d'un concept canonique, présente dans un texte, des régularités linguistiques, et il est possible d'identifier de manière algorithmique certaines d'entre elles.*

Cette hypothèse se déploie elle-même en trois sous-hypothèses, qui chacune génèrent trois opérations algorithmiques spécifiques, soit : (1) la génération des contextes d'un concept, (2) la classification de ces contextes, (3) et l'annotation catégorielle des classes.

3.1. Hypothèse 1 : l'analyse conceptuelle peut être réalisée par l'interprétation des contextes, d'une forme linguistique canonique, susceptible d'être l'expression d'un concept.

La première sous-hypothèse pose que l'analyse conceptuelle se réalise via l'exploration et l'interprétation des contextes linguistico-sémantiques d'une expression canonique susceptible d'exprimer un concept.[5]

La traduction algorithmique de cette hypothèse est réalisée par une technique très classique, de la concordance[6] et de ses variantes (Pincemin et al. 2006 ; McCarthy 2004). Sa fonction est de produire tous les contextes linguistico-sémantiques - phrase(s) ou énoncé(s) - d'un terme

[3] À titre d'exemple, étudier le concept d'AVORTEMENT dans un discours, c'est en saisir le contenu sémantique, les conditions d'énonciation, mais aussi les propositions qui lui donnent ce sens, et ce qu'on en infère, c'est-à-dire sa relation au concept de VIE, sa fonction normative, légale, culturelle, etc.

[4] Geoffrey Rockwell (2003): "Computing in the humanities has been plagued by resistance".

[5] Une expression canonique est une expression stabilisée dans un discours ou un corpus de textes. Lorsqu'on étudie un corpus de textes produit par une communauté d'interprétation, l'expression canonique exprime un concept socialement partagé dans la communauté.

[6] Formellement , une concordance est la production des séquences de mots T qui forment le contexte (normé) d'un mot pôle choisi $t_i \in T$ où T est dans l'ensemble des mots du textes $T = \{t_1 ... t_n\}$

pivot particulier, qui est considéré comme le point d'ancrage dans le corpus de textes, du concept étudié.

3.2. Hypothèse 2 : l'exploration des contextes d'une forme linguistique susceptible d'être l'expression d'un concept peut être réalisée par une classification automatique

La deuxième sous-hypothèse pose que les multiples contextes d'une expression canonique, présentent des régularités et celles-ci sont des indices des variations intensionnelles et inférentielles que le concept met en œuvre dans les textes analysés et interprétés.

L'outil utilisé pour assister informatiquement cette exploration est réalisé par une opération de classification textuelle automatique. La classification est appliquée sur les résultats de la concordance. Elle produit ainsi des classes de contexte (des mots pivots sélectionnés) de la concordance. Autrement dit, on classe les contextes linguistico-sémantiques d'un concept.[7]

3.3. Hypothèse 3 : les classes de contextes conceptuels peuvent être annotées de manière à catégoriser leur contenu signifiant

La troisième hypothèse, pose quant à elle, qu'il est possible d'associer à chacune des classes, une forme de description de son contenu signifiant. Cette description est une forme de catégorisation non formelle appelée annotation (Djioua et al. 2007 ; Le Priol et al. 2008 ; Meyers, 2005 ; Loper *et al*., 2002). Cette annotation touche les dimensions sémantiques, logiques, pragmatiques ou rhétoriques du travail conceptuel dans le corpus de textes.

4. Une expérimentation : l'utilisation de LACTAO pour l'interprétation du travail conceptuel collectif sur les ACCOMMODEMENTS RAISONNABLES dans les journaux québécois

Ces trois hypothèses se déploient autour de la chaîne de traitement LACTAO, qui peut être décrite en une méthodologie qui comprend cinq étapes relativement séquentielles : (1) le choix d'un corpus de textes et d'une ou de plusieurs expressions canoniques du concept que l'on veut étudier ; (2) la préparation du corpus ; (3) l'extraction des contextes linguistico-sémantique du (ou des) concept(s) sélectionné(s) ; (4) la classification automatique des contextes ; et (5) l'annotation catégorielle de ces classes. Nous présentons des résultats sur ces cinq étapes.

4.1. Étape 1 : Choix d'un corpus de textes et d'un concept

Dans le contexte sociopolitique québécois actuel (2007), existe un enjeu « de société » que l'on dit spécifique aux sociétés multiculturelles contemporaines confrontées à l'intégration de la diversité et au traitement des cas de discrimination envers les minorités. Cet enjeu s'actualise de plusieurs manières dans l'espace public. L'une d'entre elles est la controverse autour des « accommodements raisonnables », à propos de laquelle plusieurs communautés d'interprétations ont offert des réponses (les communautés scientifiques, la Cour Suprême, des groupes de citoyens, des partis politiques, etc.). Compte tenu de son actualité, sa pertinence sociale et sociologique, nous avons sélectionné le concept

[7] Dans notre recherche, classification est réalisée via des classifieurs mathématiques d'inspiration connexionniste L'objectif de recherche sous cette hypothèse n'est pas de valider la performance de l'algorithme, mais plutôt de voir si appliquée à une concordance, l'opération offre des résultats intéressants pour le chercheur des sciences sociales ou des humanités qui veut faire une analyse interprétative d'un ou de plusieurs concepts dans un corpus de textes.

d'ACCOMMODEMENT RAISONNABLE pour notre recherche. Nous l'étudions dans certains journaux québécois. Notre corpus est constitué de textes numérisés, formant 1 357 articles de journaux.[8]

4.2. Étapes 2 : Préparation du corpus

La préparation du corpus est une étape importante. Compte tenu du type de corpus de textes sélectionné, dans son format brut, celui-ci n'est pas immédiatement « traitable » par la chaîne LACTAO. Les documents (articles de journaux) doivent d'abord être réunis en un seul corpus clos. Ce corpus doit ensuite être formaté selon des paramètres précis (suppression des images, des liens hypertextes, correction des « coquilles », mise en format TXT, etc.).

4.3. Étape 3 : La concordance. Extraction des contextes linguistico-sémantiques du concept étudié

La troisième étape consiste en une extraction systématique des contextes linguistico-sémantiques spécifiques au concept d'ACCOMMODEMENT RAISONNABLE dans le corpus de texte. Cette extraction s'effectue avec une concordance. Elle permet de cibler précisément les traces empiriques textuelles directement liées au travail cognitif de conceptualisation sur les ACCOMMODEMENTS RAISONNABLES. Nous avons ainsi produit une concordance sur les mots pivots ACCOMMODEMENT et ACCOMMODEMENTS (Figure #1).[9]

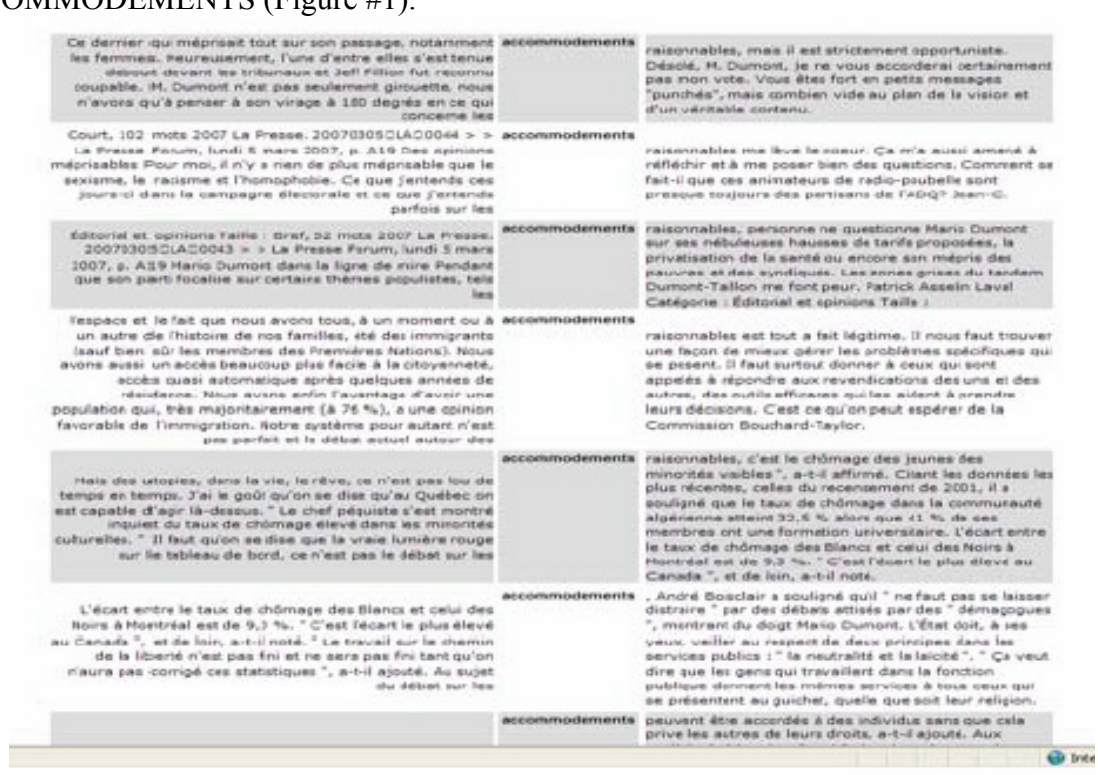

Figure #1 : Une section de concordance obtenue autour des mots pivots ACCOMMODEMENT et ACCOMMODEMENTS.

[8] Ces journaux sont La Presse, Le Devoir, Le Droit, Le Soleil, L'Actualité, Affaires plus, Commerce, Écrivains Québécois, Les Affaires, Les Affaires.com, Voir, publié entre le 6 février 1993 et le 30 juin 2007. La taille du corpus est d'environ 1 million d'occurrences, 37699 mots différents (tokens) et 2200 pages brutes de texte.

[9] Dans notre recherche, la taille des contextes est constituée de 7 phrases. Elle fut réalisé par un des modules de concordance.

Cette opération nous a donné un sous-corpus de 3 127 contextes des mots ACCOMMODEMENT et ACCOMMODMENETS, sur lequel nous avons ensuite opéré un échantillonnage aléatoire pour obtenir 1 563 contextes. Cette étape est importante, car nous obtenons ainsi un nouveau sous-corpus de textes, constitué des contextes dans lequel le concept d'ACCOMMODEMENT RAISONNABLE est spécifiquement déployé[10].

4.4. Étape 4 : Classification automatique des contextes

Vient ensuite la classification des contextes de la concordance. Il existe un très grand nombre d'outils mathématiques pour cette classification (Lebart et Salem 1994 ; Hearst 1998 ; Sebastiani 2002 ; Jain *et al*., 1999 ; Manning et Schütze 1999). Le choix d'un ou l'autre dépend de plusieurs paramètres (ampleur, différentialité, couverture, contrôle des variables, etc.). Des recherches[11] tendent à montrer cependant, que, pour les objectifs de recherches qui sont les nôtres, l'utilisation de l'un ou de l'autre affecte peu les résultats de la classification et donc de l'analyse interprétative du chercheur.

La classification est appliquée sur la concordance transformée en matrice de vecteurs [12]composée de domaines d'information (DOMIF) et d'unités d'information (UNIF) (Figure # 2).

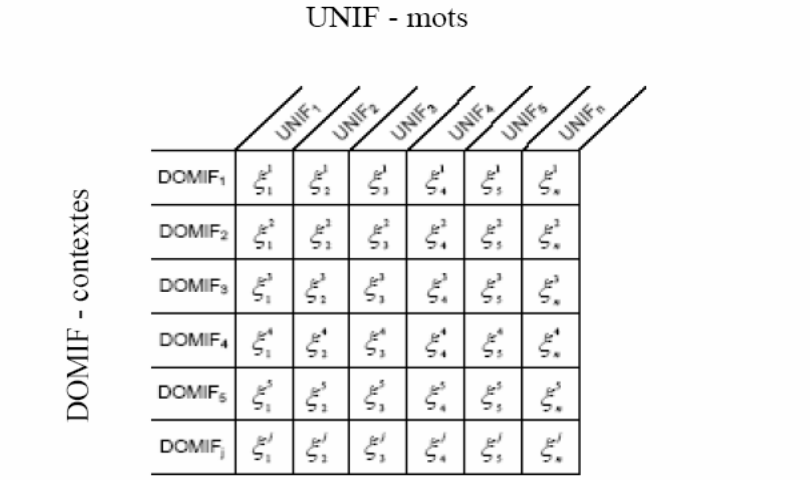

Figure # 2 : la matrice domaine d'information / unité d'information

Dans l'étude de cas présenté ici, les domaines d'information sont les 1 563 contextes des mots pivots ACCOMMODEMENT et ACCOMMODEMENTS, tandis que les unités

[10] Ce corpus est constitué de 219 000 occurrences et 15 260 mots uniques (tokens). Nous filtrons ainsi près de 80% du corpus brut : qui était composé de 985 193 occurrences alors que notre concordance est composé de 219 040 occurrences.

[11] Les multiples résultats des divers classifieurs sur des textes de référence (Reuters) varient au F score entre 0.79 et 0.87. Pour une analyse comme la nôtre, ceci peut inclure un ou deux segments mal classés sur une vingtaine de segments. Et l'essentiel du propos se retrouvera dans ceux qui ont été retenus correctement. Le bruit produit n'influence que très peu l'analyse. Voir Sebastiani, F. (2002) et Hotho et al. (2005).

[12] Selon le modèle classifieur choisi, chaque vecteur de cette matrice est constitué de la fréquence (pondérée, normalisée, etc.) de chaque unités d'information t_i pour chaque domaine d'information d (segments, contextes, pages, phrases, etc.) et défini formellement ainsi $\bar{t}_d = (tf(d,t_i),...tf(d,t_m))$

d'informations sont les 15 260 tokens de la concordance[13]. De manière générale, la classification permet d'identifier des classes d'équivalence (Jain et al., 1999 : 265) [14], c'est-à-dire des regroupements des contextes de la concordance selon des critères de similarité linguistiques (figure #3).[15]

La classification identifie les diverses classes de vecteurs ou contextes, qui expriment le travail conceptuel réalisé par l'ensemble des auteurs de notre corpus de textes : c'est-à-dire la communauté épistémique de la presse écrite québécoise. Nous avons obtenu dans notre recherche 71 classes, avec en moyenne 22 contextes par classe. En d'autres mots, ces 71 classes sont vues comme des vecteurs de la division sociale du travail cognitif de conceptualisation (De Munck 1999) sur les ACCOMMODEMENTS RAISONNABLES.

4.5. Étape 5 : l'annotation catégorielle des classes

Dans plusieurs stratégies d'analyse statistique de textes assistée par ordinateur, la production des classes et leur représentation graphique permettent d'entamer immédiatement l'interprétation. Puisque ces classifications ont été souvent appliquées sur l'ensemble du corpus et non sur un sous-corpus (une concordance) spécifique aux contextes linguistiques du concept étudié, l'analyse ne présente pas toujours les conditions de finesse essentielle dans une analyse et une interprétation conceptuelle experte des textes. Aussi, à la différence de ces approches, nous introduisons ici une étape : l'annotation. Celle-ci permet une transition plus contrôlée entre les résultats bruts des classifieurs et l'interprétation.

Dans la présente recherche, l'annotation consiste à ajouter à chaque classe, une description synthétique du contenu signifiant et de quelques propriétés du contexte d'inférence du concept étudié. Ces annotations servent d'indice pour l'interprétation experte du concept étudié. Pour le moment, cette annotation est faite à la main et à l'aide de critères statistiques, mais il est pensable qu'elle puisse être davantage assistée informatiquement (Djioua et al. 2007 ; Le Priol 2008 ; Meyers 2005 ; Loper et al., 2002)[16].

Dans cette étude, l'annotation catégorielle pour chacune des classes est assistée par une thématisions du contexte d'inférence du concept étudié. La thématisions peut être explicitée de la manière suivante :

---

[13] Sur cette matrice sont ensuite appliquées plusieurs sous-étapes implicites de lemmatisation et de filtrage des unités d'information (mot fonctionnel ou « mot vide »), qui ont pour fonction d'optimiser la classification et réduire le temps de traitement. Nous obtenons après filtrage une matrice de 1563 DOMIF et 276 UNIF. Les mots fonctionnels sont identifiés de plusieurs manières, selon des critères linguistiques et statistiques (Popping 2000; Forest 2006).

[14] En termes logiques, on peut définir la classification comme (Meunier, Remaki et Forest, 2000): un quadruplet (O, X, I, G) où O est un ensemble d'objets (o_1… o_n); X est l'ensemble des caractéristiques (x_1…. x_n) décrivant chaque objet O; I est l'ensemble des types (i_1…. i_n); G est une fonction discriminante quelconque. À partir de ces informations, une opération de classification est définie ainsi: pour tout objet de l'ensemble O, (($G (x_1 \ldots x_n)_i$)). Une telle définition, bien que très abstraite, permet de mettre en évidence que la classification n'est une fonction qui prend comme intrant des objets d'un certain type et produit un autre objet d'un même ou autre type. Et donc, que la classe produite peut être prise à son tour comme un objet indépendamment de son étiquette. Ceci permet de comprendre que la classe peut être catégorisée indépendamment de sa nomination ou de son étiquette.

[15] Dans notre étude de cas, la classification s'est faite à l'aide de l'algorithme de classification SOM de Kohonen (Kohonen 2001).

[16] Il y a là, un objet de recherche en soi, qui consiste à faire un compromis entre les avantages de la catégorisation automatique, sans inhiber la richesse herméneutique de l'analyse interprétative du chercheur.

> « [Elle] consiste à appliquer certains critères statistiques utilisés dans les domaines du repérage de l'information (pondération distribuée, tf · idf, taux d'information, entropie, etc.) (Salton, 1989) à chacune des sous-classes lexicales différenciées afin d'identifier au sein de chacune de ces sous-classes les termes les plus significatifs pouvant (suite à une évaluation de l'utilisateur) servir d'étiquette thématique pour la découverte des principaux thèmes d'un corpus. » (Forest et Meunier 2004)

Les mots utilisés en tant qu'étiquette thématique sont sélectionnés dans le lexique de chaque classe de contextes du mot ACCOMMODEMENT(S) obtenu à l'étape 4. Dans cette expérimentation, le choix des étiquettes thématiques est basé sur le critère statistique classique du tf-idf (*term frequency* * *inverse document frequency*) (Salton 1989). Dans LACTAO, le principe de cette formule peut être interprété de la manière suivante : un mot du lexique d'une classe sera d'autant meilleur pour caractérisé le contenu signifiant et le contexte d'inférence du concept étudié, s'il est très fréquent dans cette classe et rare dans les autres classes (Forest et Meunier 2004 ; Forest 2006).

Nous présentons brièvement l'annotation et l'interprétation faite sur deux classes. Nous avons retenu pour la démonstration la classe #1, constituée de 28 contextes, et la classe #15, constituée de 14 contextes.

Classe # 1 : L'accommodement raisonnable en tant que notion juridique appliquée au milieu de travail

La classe #1 est constituée de 28 contextes (figure #3). C'est à partir d'une lecture de ces contextes (segments) que le lecteur-analyste *expert* interprète le sens du concept d'ACCOMMODEMENT RAISONNABLE spécifique à cette classe.

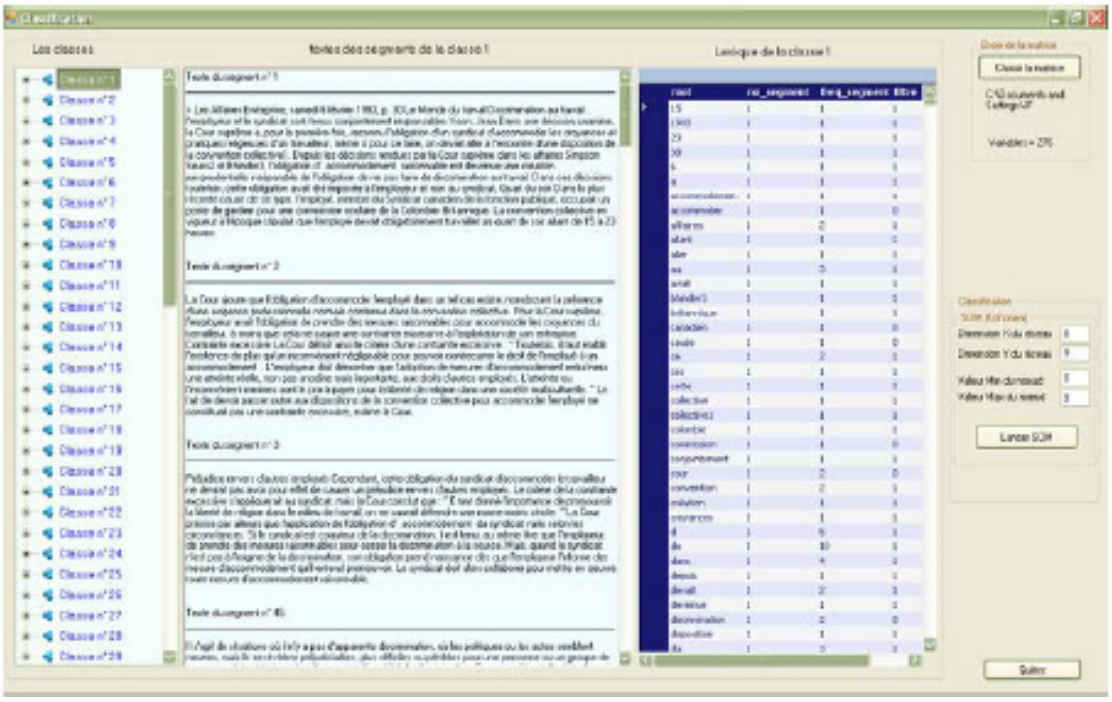

Figure #3 : Les contextes (segments) de la classe #1

Cette interprétation est assistée par une annotation catégorielle de types « thématisions de la classe ». Dans la classe #1, les étiquettes thématiques retenues pour l'annotation catégorielle sont les suivantes : EMPLOYEUR ; EMPLOYÉ ; ENTREPRISE ; CONTRAINT ;

EXCESSIVE ; OBLIGATION ; TRAVAIL ; PERSONNE ; COUR ; DISCRIMINATION ; DOIT.[17]

Une lecture assistée par ces annotations catégorielles nous permet de résumer les propriétés et le contexte d'inférence du concept d'ACCOMMODEMENT RAISONNABLE de la manière suivante : Dans la classe #1, le concept est déployé principalement dans le contexte de l'entreprise et des relations entre employeur et employé. Le concept est une notion juridique, directement issue de la Charte, pour gérer les situations de discrimination en milieu de travail. L'entreprise a l'obligation d'accorder un accommodement si cela n'entraîne pas de contrainte excessive sur les droits des autres employés et sur le bon fonctionnement, notamment financier, de l'entreprise.

Classe # 15 : L'accommodement raisonnable en tant que notion du sens commun

Pour la démonstration, on applique la même procédure à la classe #15. Les étiquettes thématiques retenues pour l'annotation catégorielle sont les suivantes : SOCIÉTÉ ; CHARTE ; PRINCIPE ; LIBERTÉ ; NÉGOCIER ; EXPRESSION ; JUGE ; SENS ; RELIGIEUX ; DROIT ; DEVOIR ; LAÏQUE.

Cela nous permet de résumer les propriétés et le contexte d'inférence du concept d'ACCOMMODEMENT RAISONNABLE de la manière suivante : Dans la classe #15, le concept est déployé dans le contexte général de la société multiculturelle laïque. Le concept conserve sont origine juridique issue de la Charte, mais devient également une expression ou un principe de sens commun ou de « gros bon sens » que les gens utilisent pour juger et négocier les droits et les devoirs face à la liberté religieuse dans une société multiculturelle laïque.

Le lecteur-analyste expert poursuit ainsi la méthode sur chaque classe obtenue à l'étape 4.[18] Cela le mène vers une interprétation finale et intégrée du concept étudié. Celle-ci met en œuvre une relecture des résultats, sous la forme de paraphrases et de commentaires. Ceux-ci sont évidemment toujours situés dans un ou plusieurs cadres théoriques. Dans le cas d'une analyse d'inspiration sociologique, cette interprétation peut appeler des explications historiques, idéologiques, structurales, culturelles, etc. Dans le cadre de cet article, nous nous sommes limités à une interprétation minimale de type « descriptive » de ce qui nous semble avoir été révélé par LACTAO dans la classe #1 et #15.

5. Conclusion

Pour les chercheurs des sciences humaines et sociales et des lettres, qui font de l'analyse conceptuelle des textes une étape importante de leur démarche scientifique, une assistance informatique sera intéressante si elle permet de diriger l'attention sur un concept précis et permettre une interprétation fine des différentes propriétés et des différents contextes d'inférence du concept étudié. Pour ces chercheurs, l'informatique ne doit pas inhiber l'imagination interprétative, mais plutôt la facilité, en devenant un levier performatif.

[17] Ce sont les mots du lexique de la classe #1 avec les scores tf-idf les plus élevé.

[18] Par exemple, dans la classe #61, le concept d'ACCOMODEMENT RAISONNABLE sert à exprimer la polarisation des politiques identitaires en deux catégories personnalisée par des leaders politiques québécois : d'un côté par Mario Dumont et de l'autre par André Boisclair. Dans la classe #46, le concept d'ACCOMODEMENT RAISONNABLE sert à mettre en évidence le clivage culturel entre d'un côté les grands centres urbains du Québec - principalement Montréal - et de l'autre ses régions, etc.

LACTAO est une méthode, qui permet avantageusement d'assister l'analyse interprétative du travail cognitif de conceptualisation déposée sous forme textuelle. Elle est une chaîne de traitement qui s'appuie sur trois opérations algorithmiques générales : l'extraction des contextes significatifs dans lesquels un concept est déployé ; la classification automatique de ces contextes ; et l'annotation catégorielle de ces classes de contextes.

Références

Alexa M. and Zuell C. (1999). *Commonalities, difference and limitations of text analysis software : the results of a review*. ZUMA arbeitsbericht, ZUMA : Mannheim.

Barry C. (1998). Choosing qualitative data analysis software : Atlas-ti and Nudist compared. *Sociological Research Online*, 3(3).

Brandom R. B. (1994). *Making it Explicit.* Cambridge, Harvard University Press.

D'Andrade R. (1995). *The development of cognitive anthropology*. Cambridge University Press, 2003.

Danis J. et Meunier J.-G. (2008). Le concept de ÉVOLUTION dans le corpus de Bergson : une analyse conceptuelle assistée par ordinateur. *Cahier du Lanci*. UQÀM.

De Munck J. (1999). *L'institution sociale de l'esprit*. Paris, PUF.

Djioua B., Desclés J.-P., Mourad G. (2007). Annotation et indexation des flux RSS par des relations discursives de citation et de rencontre : le système FluxExcom. *Analyse de texte par ordinateur, multilinguisme et applications*, 75e congrès de l'ACFAS, Trois-Rivières, Canada, 10-11 mai 2007.

Forest D. et Meunier J.-G. (2004). Classification et catégorisation automatiques : application à l'analyse thématique des données textuelles, *Proceedings of JADT 2004*.

Forest D. (2006). *Application de techniques de forage de textes de nature prédictive et exploratoire à des fins de gestion et d'analyse thématique de documents textuels non structurés*, Thèse de Doctorat, Université du Québec à Montréal.

Goody J. (1979). *La raison graphique*. Paris, Minuit.

Hearst M. A. (1998). Automated Discovery of WordNet Relations. In Christiane Fellbaum (Ed.) *WordNet : An Electronic Lexical Database*, MIT Press, p.132-152.

Hotho A., Nürnberger A, and Paass G. (2005). A Brief Survey of Text Mining. *GLDV-Journal for Computational Linguistics and Language Technologie*, 20 (1), p.19-62.

Jain A. and Flynn P. J. (1999). Data Clustering : a review. *ACM Computing Surveys*, 31(3): 264-323.

Jodelet D. (dir.) (1989). *Les représentations sociales*. Paris, PUF.

Kohonen T. (2001). *Self-Organizing Maps*. Springer.

Le Priol F., Djioua B. et Desclés J. P. (2008). *L'annotation.* Paris, Ed Hermes.

Lebart S. et Salem S. A. (1994). *Statistique textuelle*. Paris, Dunod.

Loiseau S. (2003). *Philosophical discourse from autonomy to engagement : Deleuze commentator of Spinoza*, in Fløttum et Rastier (éds), p. 36-54.

Loper E. and Bird S. (2002). *Nltk : the natural language toolkit*. CoRR, cs.CL/0205028, 2002.

Manning C. and Schutze H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge Mass., MIT Press.

McCarthy W. (2004). *Humanities Computing*. Palgrave MacMillan Blackwell Publishers.

Meunier J.-G. (2006). Le concept : de la singularité à la synthèse, *Cahier du Lanci,* UQÀM.

Meunier J.-G., Remaki L. et Forest D. (2000). Use of classifiers in computer-assisted reading and analysis of text (CARAT). Actes du colloque international CISST 1999, Las Vegas, Nevada, U.S.A.

Meunier J.-G. et Forest D. (2008). L'analyse conceptuelle assistée par ordinateur : premiers essais, (à paraître). In Le Priol F., Djioua B. et Desclés J.-P., *L'annotation*. Paris, Ed Hermes.

Meunier J.-G., Forest D. et Biskri I. (2005). Classification and Categorization in Computer Assisted Reading and Analysis of Texts. In Lefebvre C., Cohen H. (dir) *Handbook on Categorization*. Elsevier.

Meyers A. (2005). Introduction to Frontiers in Corpus Annotation II Pie. In the *Sky Proceedings of the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*. New York University Ann Arbor, p.1-4.

Pincemin B., Issac F., Chanove M. et Mathieu-Colas M. (2006). Concordanciers : thème et variations. In J.-M. VIPREY (éd.), *Proceedings of JADT 2006*, p. 773-784.

Popping R. (2000). *Computer-assisted text analysis*. London : Sage.

Rastier F. (2005). Pour une sémantique des textes théoriques. *Revue de sémantique et de pragmatique*, 17, 2005, p.151-180.

Reinert M. (1994). Quelques aspects du choix des unités d'analyse et de leur contrôle dans la méthode Alceste. In L. L. S. Bolasco and A. Salem (eds.). *Analisi Statistica dei Dati Testuali*, Vol. 1. Rome, CISU, p.19-27.

Rey G. (1983). Concepts and stereotypes. *Cognition,* Vol. 15, p.237-62.

Rockwell G. (2003). What is text analysis, really? *Literary and Linguistic Computing*, 18(2): 209-219.

Salton G. (1989). *Automatic Text Processing*. Addison-Wesley.

Sebastiani F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, Vol. 34, no 1, p.1-47.

Unsworth J. (2005). *New Methods for Humanities Research, Lyman Award Lecture 2005*, manuscrit disponible sur le Web.

Valette M. (2003). Conceptualisation and Evolution of Concepts. The example of French Linguist Gustave Guillaume, in Kjersti Fløttum et François Rastier (eds) *Academic Discourse-Multidisciplinary Approaches*, Oslo, Novus.

Wilson T. (2001). Review of Atlas-ti. *Information Research*, 6(3).