

Analyse exploratoire d'entrevues de groupe : les jeunes Français et le tabac

Claire Gélinas-Chebat¹, François Daoust², Monique Dufresne³,
Karine Gallopel⁴, Marie Éline Lebel⁵

¹Professeure, DLDL, UQAM – Montréal – Canada – chebat.claire@uqam.ca

²Informaticien au Centre ATO, doctorant U. de Franche-Comté – Besançon – France
daoust.francois@uqam.ca

³Professeure associée, DLDL, UQAM – Montréal – Canada

⁴Maître de conférences, Université de Rennes – Rennes – France

⁵Chercheure, DLDL, UQAM – Montréal – Canada

Abstract

This paper presents a methodology related to text analysis constituted from a teenager group of discussion about tobacco usage. The emphasis of the methodology point out a construct of a content analysis based on interactive process, which leads to data base building and hypothesis-testing statistical analysis. Statistical comparison in between lexical data from sub-categorization of the text associated with sociological information permit us to lead and relate our research questions to the textual analysis.

Résumé

Cet article a pour objectif de montrer, au moyen d'une analyse exploratoire d'un corpus constitué de transcriptions d'entrevues de groupe sur l'usage du tabac, une méthodologie d'analyse textuelle qui vise à construire un système de catégories de manière itérative. Des algorithmes simples de comparaison statistique entre lexiques associés à des sous-textes correspondant à nos variables sociologiques permettent d'appuyer la construction d'un système catégoriel qui établit un pont entre la problématique de recherche et nos données textuelles.

Mots-clés : analyse de textes, corpus d'entrevues de groupes, catégorisation, approche inductive et itérative, analyse lexicale.

1. Introduction

Nous présentons ici une analyse d'entrevues de groupe sur l'usage et les attitudes à l'égard de la consommation du tabac. Il s'agit d'un traitement exploratoire de transcriptions en vue de mettre en place, dans un premier temps, une méthodologie permettant le traitement textuel et itératif des données et, dans un deuxième temps, de s'assurer d'un protocole expérimental adapté à la saisie de nouvelles données dans un contexte similaire. En effet, cette recherche constitue une étape préparatoire à la saisie complémentaire de données du même type, cette recherche est le premier volet d'un projet de recherche plus vaste qui se réalisera dans un contexte français, canadien et américain.

Outre le traitement lexical, le traitement exploratoire de nos données tient compte du profil sociologique de nos sujets afin de proposer une grille catégorielle qui permet de comprendre, à travers les interventions de chacun, l'influence de messages d'avertissement sur l'usage du tabac. Les résultats de cette analyse seront exploités plus à fond dans les autres volets de cette recherche.

La première partie décrit notre problématique et aborde la question de l'attitude des adolescents à l'égard du tabagisme et des messages dissuasifs. Nous enchaînons sur les caractéristiques de notre corpus à la section suivante. La quatrième partie expose notre méthodologie interactive d'analyse textuelle. Enfin, la dernière partie présente des extraits du tableau de bord de la recherche montrant à l'œuvre notre démarche dans ses phases inductive et hypothético-déductive.

2. Problématique et questions de recherche

Cette recherche tente de comprendre la portée des messages antitabac chez les adolescents avec l'objectif à long terme de réduire significativement leur consommation de cigarettes. Malgré des efforts marqués pour dissuader les adultes à fumer et convaincre les jeunes à ne pas commencer à fumer, la consommation des produits du tabac reste très importante (Santé Canada *et al.*, 1999). En France, près d'un jeune sur deux fume.

Les adolescents minimisent les risques de l'usage des produits dangereux et tendent à sous-estimer les dangers de l'usage du tabac (Leventhal *et al.*, 1987). Quel discours doit-on tenir dans ces messages d'avertissement si on veut produire des messages efficaces ? Une étude comparative de différentes recherches empiriques sur les effets des menaces dans le domaine des mises en garde sur la santé montre que plus le message suscite des émotions de peur, plus les effets sur l'attitude, l'intention et les changements de comportements sont grands. De même, plus la sévérité du message est forte, plus l'attitude, l'intention et les comportements changent (Witte et Allen, 2000). Mais le goût du risque, par exemple des sports extrêmes et l'attrait du « fruit défendu » (Parker-Pope, 1997) n'a-t-il pas l'effet contraire à l'effet désiré ?

Dans la démarche exploratoire, nous avons traité les données recueillies dans le cadre d'une discussion ; des groupes de plusieurs participants discutaient sur le thème de la cigarette, puis ils étaient exposés à différents messages d'avertissement avec des menaces graduées. Chaque groupe était exposé à un seul message d'avertissement. À cette étape-ci de la recherche, nous nous sommes mis à l'écoute des jeunes exposés à ces différents messages dans le but de reprendre les arguments évoqués ultérieurement afin de construire dans une deuxième phase expérimentale des messages persuasifs dont nous mesurerons expérimentalement les effets.

3. Un corpus de groupes de discussion

Ce corpus comprend neuf entrevues sur le tabagisme chez les jeunes et leur perception de la publicité antitabac. Elles ont été réalisées à Rennes en 2000 auprès de 48 jeunes Français qui, pour la plupart, fréquentent une institution scolaire et qui sont âgés de 15 à 25 ans. Chacune des séances réunit 5-6 jeunes (fumeurs ou non-fumeurs, hommes et femmes) et un intervenant, et se divise en deux parties. La première partie se déroule après que l'intervenant a posé quelques questions pour amorcer la discussion et la seconde se caractérise par l'introduction d'une brochure. Il existe différentes versions de la brochure selon deux paramètres : les effets du tabagisme sur la santé et les solutions pour arrêter de fumer. Ces deux paramètres se présentent comme suit, trois niveaux de menace (faible, moyen et fort), et deux niveaux de solution (faible et fort).

Les entrevues ont été enregistrées sur bandes audio et retranscrites en format Word. Au début de chaque transcription, les données sociologiques des personnes qui participent à l'entrevue sont précisées : âge, sexe, fumeur/non-fumeur. Nous avons effectué l'analyse du corpus au moyen du logiciel SATO (Daoust, 1996). Les annotations éditiques ont été remplacées par un balisage symbolique conforme à la syntaxe de SATO. Un astérisque introduit les balises,

aussi appelées *propriétés*, et celles de notre corpus sont les suivantes : *locuteur, *sexe, *fumeur, *page et *thème. En voici un exemple :

*page=gallo02/11

*thème=brochure (...) *locuteur=s36 *fumeur=non *sexe=ho Bah, la brochure là, elle nous présente ce qui nous attend si on fume. Mais c'est très... quoi, moi j'ai lu ça, mais je ne sais pas je ne suis pas fumeur, donc je ne ressens peut-être pas ça de la même façon. À la limite on passe dessus comme ça, ça apporte quelques chiffres.

4. Une construction itérative des catégories

Nous avons utilisé une démarche exploratoire inspirée de l'approche adoptée pour le corpus *Message d'amour* (Daoust, 1999) et qui vise à *faire parler les données*. Le principe de base de la démarche consiste à comparer, avec des indices statistiques simples, les lexiques associés à des sous-textes découpés d'après nos variables sociologiques. Cette comparaison sera reprise de façon itérative de façon à s'appuyer sur le lexique brut pour construire un lexique catégorisé reflétant les points d'ancrage de notre chaîne interprétative.

Le découpage du texte et la constitution des lexiques associés procèdent comme suit. Les balises (*propriétés*) introduites dans notre corpus permettent de segmenter le corpus en opposant, par exemple, l'ensemble des interventions *avant* et *après* la présentation du document publicitaire. De la même façon, on peut découper le corpus entre, d'une part, les interventions des hommes et, d'autre part, celles des femmes, excluant les interventions des modérateurs. Ces balises étant indépendantes, elles peuvent être combinées à loisir lors de l'exploration du corpus. Ainsi, on pourrait comparer les interventions des hommes seulement, ou des femmes seulement, *avant* et *après* la présentation de la brochure pour voir si la réaction à la brochure dépend du sexe des sujets.

Pour comparer nos lexiques, nous utilisons un *algorithme de distance lexicale* basé sur la *distance du Chi2*. La mesure évalue l'écart dans l'utilisation d'un vocabulaire donné entre deux sous-ensembles du corpus. Les formes lexicales sont triées par ordre décroissant de contribution à la mesure de distance, ce qui permet d'identifier, par ordre d'importance, les spécificités de chaque sous-texte. L'algorithme peut être appliqué aux formes lexicales elles-mêmes ou aux valeurs de propriétés correspondant à notre catégorisation lexicale. Ici, l'approche est essentiellement dichotomique : on compare un sous-texte à un autre, via leur lexique respectif. On peut aussi avoir recours à un *algorithme de participation* qui calculera les moyennes normalisées d'un ensemble de formes lexicales, correspondant généralement à une catégorie lexicale, pour chacun des sous-textes constitués en cours d'analyse.

Notre démarche exploratoire est fondée sur un va et vient interactif entre ce que nous révèle l'analyse lexicale et les contextes d'utilisation des mots mis en évidence par les algorithmes de distance et de participation. Dans ce premier temps de l'analyse, nous privilégions une approche univariée afin de mieux saisir la spécificité de la stratification induite par chacune de nos variables sociologiques. Dans un deuxième temps, on peut, avec les mêmes outils, comparer des sous-textes faisant appel à plus d'une variable, comme dans l'exemple ci-haut tenant compte du sexe et de l'introduction de la brochure.

Au premier niveau de l'analyse, nous travaillons sur les données brutes, c'est-à-dire les formes lexicales elles-mêmes. On se donne ainsi la possibilité de voir apparaître des différenciations portées par la morphologie des mots en termes de nombre, genre, personne, temps. On s'intéresse tout autant, sinon davantage, à des marqueurs d'énonciation, comme les pronoms personnels, les marqueurs phatiques, les marques de la négation, de l'interrogation, les verbes

épistémiques (croire, penser...), etc. qu'aux termes pleins. C'est en s'appuyant sur l'analyse lexicale des données brutes que nous élaborerons nos grilles catégorielles.

Le retour constant aux énoncés, ne serait-ce que par un parcours rapide des contextes courts de type KWIC (*key words in context*), est cependant essentiel pour ébaucher nos hypothèses sur le fonctionnement du discours et le positionnement des locuteurs d'après leur profil social. Ce va et vient entre l'analyse lexicale et les énoncés permet en effet d'inscrire les unités lexicales dans des systèmes de catégories sémantiques et énonciatives susceptibles de traduire, dans le discours même, ce que l'on cherche à comprendre, à savoir ici l'attitude des jeunes par rapport au tabagisme et l'influence de publicités dissuasives. La catégorisation vise donc à établir le pont entre la problématique de recherche et nos données textuelles. Elle correspond un peu à la procédure de codage de l'analyse qualitative à cette différence qu'elle s'appuie sur des procédures d'analyse lexicale qui permettent de tenir compte de l'ensemble des données et sur l'examen de phénomènes discursifs difficilement repérables par une simple lecture linéaire.

La reprise des analyses univariées et multivariées sur les données catégorisées s'inscrit dans les procédures de validation de nos hypothèses interprétatives. Notre première approche, inductive, qui implique à la fois une sensibilité aux procédés linguistiques et à la problématique de la recherche, sera donc relayée par une approche davantage hypothético-déductive.

5. Tableaux... d'une exploration

L'analyse de distance permet de déterminer le vocabulaire qui caractérise un sous-texte, c'est-à-dire les formes qui marquent davantage l'originalité du vocabulaire d'une partie du corpus par rapport à l'autre. Dans le tableau qui suit (tableau I), les mots qui caractérisent le plus le discours avant l'introduction de la brochure sont suivis d'un astérisque. Les mots sans astérisque caractérisent davantage les propos tenus après l'introduction de la brochure.

	*				
Fréqtot	A	B	explique	cumul	
0.07	0.14	0.02	0.44	0.44	clair *
0.23	0.38	0.18	0.40	0.84	aussi *
0.05	0.11	0.02	0.31	1.15	plaisir *
0.06	0.11	0.02	0.31	1.46	dépendance *
0.02	0.00	0.05	0.28	1.75	témoignage
0.09	0.04	0.15	0.28	2.02	"
0.01	0.03	0.00	0.26	2.28	3ème *
0.02	0.05	0.00	0.25	2.54	doigts *
0.06	0.01	0.09	0.24	2.78	risques
0.02	0.05	0.00	0.24	3.02	primaire *
0.37	0.45	0.25	0.24	3.25	ils *
0.59	0.62	0.87	0.23	3.49	j'
0.03	0.01	0.06	0.23	3.72	concret
0.01	0.00	0.04	0.23	3.95	cinq
0.09	0.13	0.04	0.22	4.17	santé *
0.02	0.00	0.04	0.21	4.38	solution
0.02	0.04	0.00	0.20	4.58	appelle *
0.02	0.00	0.05	0.20	4.78	chiffres
0.03	0.01	0.06	0.20	4.98	routière

0.01	0.03	0.00	0.19	5.17	choqué *
0.01	0.03	0.00	0.19	5.37	influencé *
0.01	0.03	0.00	0.19	5.56	dents *
0.15	0.10	0.21	0.19	5.74	elle
0.36	0.36	0.53	0.18	5.93	!
0.01	0.00	0.03	0.18	6.11	morts
0.03	0.02	0.06	0.18	6.29	y'
0.02	0.04	0.00	0.18	6.47	dérange *
0.07	0.04	0.11	0.18	6.65	Cela
0.28	0.17	0.32	0.18	6.83	Te
0.02	0.00	0.04	0.18	7.00	Image
0.02	0.00	0.04	0.18	7.18	Provoque
0.02	0.04	0.00	0.17	7.35	odeur *
0.04	0.09	0.03	0.17	7.52	effectivement *
0.02	0.04	0.00	0.17	7.69	jaunes *
0.16	0.17	0.06	0.17	7.86	toi *
0.06	0.10	0.03	0.17	8.03	niveau *
0.01	0.03	0.00	0.17	8.20	publicitaire *
0.02	0.00	0.04	0.17	8.36	Long

Tableau 1. Analyse de distance sur les formes lexicales brutes avant/après l'introduction de la brochure

Si on s'attarde aux items lexicaux pleins, c'est-à-dire aux noms, adjectifs et aux verbes, il semble que les mots qui décrivent l'apparence physique et la santé en général sont ceux qui caractérisent le plus vocabulaire avant l'introduction de la brochure *clair*, *doigts*, *dents*, *santé* sans oublier la notion de *plaisir* et de *dépendance* ; après l'introduction de la brochure, il est remarquable de constater que les mots *témoignage*, *concret*, *solution*, *chiffres*, *mort* sont ceux qui apparaissent en tête de liste. Il appert que les deux sous textes ne font pas ressortir les effets du tabac dans les mêmes termes. Avant on parle de plaisir et des effets néfastes sur la santé et particulièrement sur l'apparence physique, les dents et les doigts jaunes et sur la dépendance. Après, les effets font toujours partie du discours, mais alors non plus en termes de plaisir, mais en termes de risque et de mort. Notons par ailleurs la présence importante du pronom *j'* après l'introduction de la brochure. Ceci pourrait suggérer que la brochure provoque une plus grande implication personnelle de nos sujets.

Pour valider ces observations, la première stratégie de vérification consiste à parcourir rapidement les contextes. Ainsi, nous avons constaté que le mot *clair* n'a rien à voir avec l'apparence, mais est plutôt utilisé comme marque évaluative : *C'est clair*, *c'est évident*.

Nous avons poursuivi nos travaux en nous donnant une procédure de vérification des hypothèses d'interprétation construites à partir des premiers résultats obtenus lors de l'analyse de distance sur les unités lexicales brutes. C'est ici qu'entre en jeu la catégorisation des unités lexicales. Il s'agit de déterminer avec plus de raffinement les sujets traités au cours des discussions, toujours dans la perspective de déterminer si l'introduction de la brochure produit des changements dans le discours des participants.

Nous avons donc introduit une propriété lexicale que nous avons nommée *sujet*. Cette propriété englobe en fait plusieurs grilles d'analyse que nous aurions pu regrouper dans des propriétés différentes. Mais dans l'idée de procéder de façon itérative par raffinements successifs de nos procédures, il était suffisant, à ce stade-ci, de n'avoir qu'un système de catégories dont certaines peuvent déjà faire l'objet d'une structuration plus fine. Ainsi, seront rajoutées à

notre grille cinq catégories (*Soc-X*) associées aux formes lexicales décrivant l'environnement social. Pourquoi ? Nous avons remarqué au cours de l'analyse de distance que les items lexicaux qui faisaient référence à l'aspect social du tabagisme, c'est-à-dire de ses conséquences sur les rapports sociaux des jeunes, ressortaient beaucoup. On a donc établi en suffixe une échelle décrivant le niveau de l'environnement social, allant du plus intime au plus général : *soc-je*, *soc-ami*, *soc-famille*, *soc-jeune*, *soc-gens*.

Voici le descriptif de la propriété *sujet*. On a déterminé 28 *sujets* : *apparence*, *arrêt*, *négation*, *concret*, *danger*, *dépendance*, *soc-je*, *maladie*, *mort*, *plaisir*, *publicité*, *tabac*, *nicotine*, *drogue*, *interdiction*, *fumeur*, *soc-ami*, *soc-famille*, *soc-gens*, *liberté*, *envie*, *conscience*, *volonté*, *soc-jeune*, *coûts*, *début*, *santé*, *éducation*, *prévention*.

La procédure de catégorisation procède des mots caractéristiques révélés par l'algorithme de distance, vers l'ensemble du vocabulaire. Après ces mots, nous avons examiné de façon systématique les mots fréquents et complété la catégorisation en examinant le lexique trié par ordre alphabétique pour catégoriser les variantes flexionnelles pertinentes. Pour confirmer nos intuitions, nous reprenons les analyses statistiques sur les valeurs de la propriété *sujet* comme l'illustre le tableau II qui suit.

	*				
Fréqtot	A	B	explique	cumul	
0.21	0.43	0.11	31.23	31.23	apparence *
0.09	0.02	0.16	13.85	45.08	Concret
0.08	0.14	0.05	6.75	51.83	plaisir *
0.13	0.21	0.10	6.63	58.46	dépendance *
0.14	0.19	0.08	5.64	64.10	santé *
0.11	0.17	0.08	5.39	69.49	éducation *
0.18	0.11	0.22	5.12	74.61	Volonté
0.10	0.08	0.17	4.75	79.36	Mort
1.95	2.19	1.82	4.53	83.89	tabac *
0.05	0.10	0.05	3.26	87.15	soc-ami *
0.17	0.25	0.16	3.12	90.27	coûts *
0.32	0.28	0.40	2.81	93.09	Maladie
0.75	0.59	0.72	1.44	94.53	Publicité
0.21	0.26	0.20	1.37	95.90	soc-famille *
0.11	0.14	0.11	0.84	96.74	drogue *
0.20	0.22	0.17	0.82	97.55	liberté *
0.74	0.69	0.78	0.67	98.23	soc-gens
0.17	0.14	0.18	0.66	98.89	Envie
0.31	0.29	0.24	0.54	99.44	soc-jeune *
0.05	0.08	0.06	0.19	99.62	nicotine *
0.63	0.60	0.64	0.18	99.81	Arrêt
0.08	0.07	0.08	0.09	99.90	Conscience
0.24	0.17	0.19	0.03	99.93	Danger
0.22	0.17	0.16	0.03	99.96	début *
0.13	0.11	0.12	0.02	99.98	Prévention
2.28	2.48	2.50	0.01	99.99	Négation
0.48	0.44	0.45	0.01	100.00	Fumeur
2.14	2.54	2.53	0.00	100.00	soc-je *

Tableau 2. Analyse de distance sur les formes de la catégorie sujet avant/après l'introduction de la brochure

La notion d'*apparence* se confirme. Les sujets, avant la brochure, abordent les effets superficiels du tabagisme, à savoir, la couleur des dents et des doigts, le teint, l'odeur des vêtements et des cheveux... La notion de plaisir ressort aussi comme thème avant l'introduction de la brochure, ainsi que les notions de dépendance, santé et éducation.

Après l'introduction de la brochure, la catégorie *concret* (impact et solutions) ressort. Nous voyons aussi émerger les notions de *volonté*, *mort* et *maladie*. D'autre part, nous constatons que notre hypothèse sur les pronoms personnels de la première personne (*soc-je*) ne se confirme pas. L'écart observé avant et après la brochure était spécifique à la forme *j'*.

Nous affinons l'analyse en comparant les interventions avant et après l'introduction de la brochure selon le profil sociologique des sujets. Ainsi, en comparant les données qui suivent (tableau III), il est remarquable de constater que ce sont les non-fumeurs qui semblent le plus touchés par la brochure comme en témoigne la dominance des thèmes relatifs aux effets négatifs du tabagisme : *maladie* et *mort*.

Comparaison « avant – après » pour les fumeurs						Comparaison « avant – après » pour les non-fumeurs					
Mode propriété sujet						Mode propriété sujet					
Fréqtot	Afu	Bfu	explique	cumul		Fréqtot	Anf	Bnf	explique	cumul	
0.21	0.47	0.11	32.55	32.55	apparence*	0.75	0.42	0.99	15.22	15.22	publicité
0.09	0.03	0.19	18.05	50.60	concret	0.21	0.38	0.10	12.52	27.74	apparence*
0.11	0.20	0.06	10.15	60.75	éducation*	0.05	0.14	0.02	10.64	38.38	soc-ami*
0.13	0.20	0.08	6.26	67.01	dépendance*	0.17	0.37	0.15	9.73	48.11	coûts*
0.18	0.07	0.21	5.79	72.79	volonté	0.14	0.20	0.03	8.21	56.32	santé*
0.08	0.15	0.06	4.85	77.64	plaisir*	0.32	0.29	0.54	6.94	63.26	maladie
0.48	0.35	0.53	3.96	81.60	fumeur	0.10	0.06	0.19	5.96	69.22	mort
0.17	0.15	0.25	3.77	85.37	envie	0.08	0.14	0.04	4.51	73.72	plaisir*
0.75	0.71	0.51	2.70	88.07	publicité*	1.95	2.32	1.82	4.38	78.11	tabac*
1.95	2.09	1.81	2.23	90.30	tabac*	0.09	0.01	0.11	4.01	82.11	concret
0.21	0.28	0.20	1.79	92.08	soc-famille*	0.48	0.57	0.34	4.00	86.12	fumeur*
0.10	0.09	0.15	1.74	93.83	mort	0.13	0.24	0.13	3.15	89.27	dépendance*
0.14	0.18	0.12	1.60	95.43	santé*	0.11	0.19	0.11	2.32	91.59	drogue*
0.22	0.18	0.12	0.99	96.42	début*	0.20	0.20	0.13	1.09	92.68	liberté*
0.63	0.57	0.67	0.98	97.41	arrêt	0.05	0.08	0.04	1.05	93.73	nicotine *
2.14	2.74	2.93	0.91	98.32	soc-je	2.14	2.25	2.01	0.93	94.66	soc-je *

Tableau 3. Analyse de distance avant/après pour les fumeurs et les non-fumeurs

Une première analyse des résultats obtenus nous a amenés à conclure que les hommes semblaient plus interpellés par l'introduction de la brochure que les femmes. Ce résultat confirme-t-il les autres analyses sur les effets des campagnes antitabac, à savoir que les femmes sont moins touchées que les hommes ? Une réponse positive, à cette étape de l'analyse, serait prématurée.

Une autre façon de visualiser les résultats nous est donnée par l'analyseur PARTICIPATION. Pour une catégorie donnée, l'analyseur calcule sa fréquence relative dans les divers sous-tex-

tes, ce qui permet de voir si les variables sociologiques ont une influence sur l'utilisation des mots catégorisés. Les tableaux IV et V illustrent la distribution des catégories *apparence* et *mort* dans le corpus complet et divers sous-textes. **A** et **B** désignent *avant* et *après* la brochure. Nous avons aussi les particules **fu** et **nf** pour *fumeur* et *non-fumeur*, ainsi que **ho** et **fe** pour *homme* et *femme*.

Propriété	Couverture	Lexèmes	Occurrences	Cote Z
Fréqtot	78703/78703 (100.00%)	37/3985 (0.93%)	168/78703 (0.21%)	0.00
A	23544/78703 (29.91%)	30/2087 (1.44%)	101/23544 (0.43%)	7.17
B	28074/78703 (35.67%)	18/2351 (0.77%)	30/28074 (0.11%)	-3.87
Afu	13758/78703 (17.48%)	24/1580 (1.52%)	64/13758 (0.47%)	6.40
Bfu	15923/78703 (20.23%)	13/1749 (0.74%)	18/15923 (0.11%)	-2.75
Anf	9786/78703 (12.43%)	19/1240 (1.53%)	37/9786 (0.38%)	3.53
Bnf	11898/78703 (15.12%)	8/1425 (0.56%)	12/11898 (0.10%)	-2.66
Aho	14468/78703 (18.38%)	16/163 (4 0.98%)	44/14468 (0.30%)	2.36
Bho	16010/78703 (20.34%)	11/1797 (0.61%)	19/16010 (0.12%)	-2.60
Afe	9076/78703 (11.53%)	24/1153 (2.08%)	57/9076 (0.63%)	8.56
Bfe	11811/78703 (15.01%)	9/1379 (0.65%)	11/11811 (0.09%)	-2.83

Tableau 4. Analyseur PARTICIPATION (sujet=*apparence*)

Propriété	Couverture	Lexèmes	Occurrences	Cote Z
Fréqtot	78703/78703 (100.00%)	9/3985 (0.23%)	80/78703 (0.10%)	0.00
A	23544/78703 (29.91%)	4/2087 (0.19%)	19/23544 (0.8%)	-1.01
B	28074/7870335 (67%)	6/2351 (0.26%)	47/28074 (0.17%)	3.46
Afu	13758/78703 (17.48%)	4/1580 (0.25%)	13/13758 (0.09%)	-0.26
Bfu	15923/78703 (20.23%)	6/17490.(34%)	24/15923 (0.15%)	1.94
Anf	9786/7870312.(43%)	2/1240 (0.16%)	6/9786 (0.06%)	-1.25
Bnf	11898/78703 (15.12%)	3/1425 (0.21%)	23/11898 (0.19%)	3.14
Aho	14468/78703 (18.38%)	4/1634 (0.24%)	8/14468 (0.06%)	-1.75
Bho	16010/78703 (20.34%)	4/1797 (0.22 %)	21/16010 (0.13%)	1.17
Afe	9076/78703 (11.53%)	2/1153 (0.17%)	11/9076 (0.12%)	0.58
Bfe	11811/78703 (15.01%)	5/1379 (0.36%)	26/1181 (0.22%)	4.04

Tableau 5. Analyseur PARTICIPATION (sujet=*mort*)

On a repris cette procédure de comparaison des interventions avant et après l'introduction de la brochure, mais en distinguant selon le niveau de *menaces* contenu dans la brochure. On a noté que la présence d'une menace forte provoque un débat sur la liberté et la contrainte. Si on tient compte de la variable *fumeur*, cette préoccupation en vient à occuper le premier rang. Il y a là sans doute des leçons à tirer pour les campagnes antitabac.

Ces quelques exemples, extraits d'une démarche qui sera reprise et développée sur un corpus québécois en cours de constitution, visent d'abord à montrer une approche permettant de construire un protocole de constitution et d'analyse de corpus qui soit à la fois transparent et respectueux de la spécificité du contexte d'énonciation, ici des échanges oraux analysés sous forme de transcriptions. Mentionnons notamment que cette analyse exploratoire a mené à certaines modifications du protocole d'entrevue. Notamment, nous voulions limiter l'introduction par l'intervenant de sujets ou de thèmes qui n'auraient pas été abordés par les jeunes eux-mêmes. Notre protocole expérimental de la deuxième phase de la recherche utilisait des formulations beaucoup moins suggestives qu'au moment de la première phase d'expérimentation. Par exemple, la question *Y a-t-il des conséquences négatives à fumer ?* a été reformulée par *Que pensez-vous que ça apporte de fumer la cigarette ?* Notre première phase de recherche nous a permis de mettre en place une procédure d'analyse textuelle que nous comptons appliquer à nouveau.

Pour conclure, la démarche d'analyse lexicale de nos données est une démarche itérative qui combine l'approche inductive, souvent associée aux méthodes qualitatives, l'utilisation d'outils simples de statistique lexicale, et une approche plus sensible à la pragmatique textuelle. Ce traitement textuel des données a aussi l'avantage de produire des données qualifiées pouvant être soumises à des algorithmes statistiques multivariés susceptibles de mesurer les corrélations entre les variables sociologiques et les différentes modalités catégorielles construites au cours du processus d'analyse textuelle.

Références

- Daoust Fr. (1996). *SATO 4, Manuel de référence*. Centre ATO, UQAM, Montréal.
- Daoust Fr. (1999). *Corpus Message d'amour : analyse exploratoire*. Centre ATO, Montréal, <http://www.ling.uqam.ca/sato/analyses/amour1.html>.
- Gilmore J. (2000). *Report on Smoking Prevalence in Canada, 1985 to 1999, Statistiques Canada*. Catalogue 82F00077XIE, Ottawa : Ministère de l'Industrie et du commerce du Canada.
- Parker-Pope T. (1997). Danger: Warning Labels May Backfire. *Wall Street Journal* (April 28), B1, B8.
- Santé Canada, Statistics Canada, Canadian Institute for Health Information (1999). *Statistical Report on the Health of Canadians*. Ottawa. Minister of Public Works and Government Services Canada.
- Snyder L.B. et Blood D.J. (1992). Caution: Alcohol and The Surgeon General's Alcohol Warnings May Have Adverse Effects on Young Adults. *Journal of Applied Communication Research*, vol. (20/1): 37-53.
- Witte K. et Allen M. (2000). A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns. *Health Education & Behavior*, vol. (27/5) : 591-616.