Nouvelles perspectives en sciences sociales
Revue internationale de systémique complexe et d'études relationnelles

Réflexion méthodologique sur l'usage des logiciels Modalisa et Iramuteq pour l'étude d'un corpus de presse sur l'anorexie mentale

Methodological reflection on the use of Modalisa and Iramuteq Software for the study of a corpus of newspapers about anorexia nervosa

Audrey Arnoult

Volume 11, Number 1, November 2015

Sur le thème de l'analyse de données textuelles informatisée

URI: https://id.erudit.org/iderudit/1035939ar
DOI: https://doi.org/10.7202/1035939ar

See table of contents

Publisher(s)

Prise de parole

ISSN

1712-8307 (print)
1918-7475 (digital)

Explore this journal

Cite this article

Arnoult, A. (2015). Réflexion méthodologique sur l'usage des logiciels Modalisa et Iramuteq pour l'étude d'un corpus de presse sur l'anorexie mentale. *Nouvelles perspectives en sciences sociales*, *11*(1), 285–323. https://doi.org/10.7202/1035939ar

Article abstract

Anorexia nervosa is a complex multi-factorial disease currently considered as a public health problem by the medical profession. However, media discourses on this subject are relatively recent. This article aims to understand the characteristics of the media coverage of this pathology and the representations produced by the media of this disorder during adolescence while showing how the use of software for automated analysis of textual data can be helpful. For this purpose, we conducted a quantitative and content analysis of a corpus of 131 articles, published between 1995 and 2009, in several French daily newspapers, with the Modalisa software. Then, we used the Iramuteq software to identify lexical worlds structuring the discourses, based on a second more limited corpus.

Réflexion méthodologique sur l'usage des logiciels Modalisa et Iramuteq pour l'étude d'un corpus de presse sur l'anorexie mentale

Audrey Arnoult
Laboratoire Elico (Équipe de recherche de Lyon en Sciences de l'Information et de la Communication), Lyon

Introduction

L'anorexie mentale est attestée dès le Moyen Âge dans les écrits religieux mais il faut attendre les travaux du psychiatre Charles Lasègue[1], à la fin du XIX[e], pour qu'elle soit reconnue comme une pathologie. Les recherches de ce médecin ouvrent une période riche en questionnement sur l'étiologie de cette maladie et les thérapies susceptibles de guérir les jeunes filles qui en souffrent. L'anorexie mentale est aujourd'hui reconnue comme une maladie polyfactorielle aux conséquences graves. Elle toucherait environ 0,5 % des adolescentes et 0,03 % des adolescents entre 12 et 17 ans[2]. Bien que l'anorexie soit considérée comme un problème de santé publique par le corps médical depuis

1 Son article « De l'anorexie hystérique », publié dans les *Archives générales de médecine* en 1873, pose les bases de la définition de l'anorexie mentale; Charles Lasègue, « De l'anorexie hystérique », *Journal français de psychiatrie*, n° 32, 2009, p. 3-8.

2 Inserm, 2014, http://www.inserm.fr/thematiques/neurosciences-sciences-cognitives-neurologie-psychiatrie/dossiers-d-information/anorexie-mentale, site consulté le 25 mars 2015.

plusieurs années, les discours médiatiques sur cette maladie sont relativement récents[3].

Notre contribution s'inscrit dans le champ des sciences de l'information et de la communication. Dans une perspective constructiviste, nous considérons les discours médiatiques comme le fruit d'une co-construction et non comme le reflet d'une réalité extérieure que les journalistes se contenteraient de recueillir. Les journalistes sont des acteurs sociaux qui mettent « en visibilité » et « en lisibilité » le monde social[4]. Leurs récits s'inscrivent dans une discursivité sociale dont ils se nourrissent et qu'ils viennent à leur tour entretenir[5]. Un discours médiatique n'est donc pas neutre mais porteur d'enjeux sociopolitiques que l'analyse des discours – ici, par le biais de logiciels – vise à repérer.

Nos questions sont les suivantes : comment se caractérise la couverture médiatique de l'anorexie mentale et quelles représentations construisent les médias de ce trouble psychique, considéré par le corps médical comme un problème de santé publique? Méthodologiquement, comment des logiciels d'analyse automatisée de discours peuvent nous permettre de saisir le sens que les discours médiatiques donnent à ce sujet? Plus précisément, quels sont les avantages de l'usage d'un logiciel d'« analyse manuelle » comparé à un logiciel d'« analyse automatique[6] », et inversement? Comment les procédures statistiques, propres à chacun de ces logiciels, permettent d'appréhender les représentations médiatiques de l'anorexie mentale? Enfin, quelle est la place du chercheur dans ce type d'analyse? N'est-t-il pas « contraint » par l'outil informatique?

3 Décès de plusieurs mannequins anorexiques (été et automne 2006), proposition de loi sanctionnant l'apologie de l'anorexie (2008), etc.

4 Bernard Delforce, « La responsabilité sociale du journaliste : donner du sens », *Les Cahiers du journalisme*, n° 2, 1996, p. 28.

5 Bernard Delforce, « Discursivité sociale/discours sociaux : penser les enjeux sociaux de l'information », dans Aurélie Tavernier *et al.* (dir.), *Figures sociales des discours. Le "discours social" en perspectives*, Lille, Éditions des Presses de l'Université Charles-de-Gaulle Lille 3, 2010, p. 67.

6 Nous empruntons cette distinction à Normand Roy et Roseline Garon, « Étude comparative des logiciels d'aide à l'analyse de données qualitatives : de l'approche automatique à l'approche manuelle », *Recherches qualitatives*, vol. 32, n° 1, p. 156-157.

Pour répondre à nos questionnements, nous procédons en deux temps. Un corpus de 131 articles publiés par cinq quotidiens nationaux (*Le Monde*, *La Croix*, *Le Figaro*, *Libération* et *L'Humanité*), entre 1995 et 2009, est analysé avec Modalisa (logiciel de traitement d'entretiens et d'enquêtes). À partir des résultats obtenus, nous constituons un second corpus de 39 articles[7] qui est analysé avec Iramuteq (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires). Dans un premier temps, nous exposons les fonctionnalités des deux logiciels utilisés et ce, de façon comparative. La seconde partie de notre contribution est consacrée à la présentation des résultats tout en discutant l'apport et les spécificités de Modalisa et d'Iramuteq. Précisons que notre posture est celle d'une jeune chercheuse qui n'est ni informaticienne ni statisticienne et qui souhaite intégrer les logiciels d'analyse automatisée de discours à ses outils méthodologiques. Notre analyse est donc aussi le fruit d'une posture réflexive sur la façon dont un chercheur peut investir le champ de la statistique textuelle et s'approprier certains outils méthodologiques à un moment donné de son parcours intellectuel.

1. Les principes de l'« analyse manuelle » et de l'« analyse automatique » de discours de presse

L'analyse de données textuelles nécessite de travailler sur un corpus clos[8]. Nous présentons donc les critères qui ont guidé la constitution de notre corpus puis nous explicitons les fonctionnalités et les spécificités des deux logiciels utilisés. Nous inscrivons cette description dans une réflexion plus large sur le rôle et la place du chercheur dans le processus de l'analyse automatisée de

7 Nous revenons plus loin sur la constitution de ce second corpus.

8 En effet, comme le rappelle Damon Mayaffre, en logométrie, « Le corpus est [...] un objet physique stable, entré manuellement ou de manière semi-automatique en machine. [...] Les corpus logométriques sont clos. Les traitements informatiques tournent nécessairement sur des ensembles arrêtés » (Damon Mayaffre, « Analyses lexicologiques et rhétoriques du discours », 1er-11 décembre 2009, Alexandrie, http://eprints.aidenligne-francais-universite.auf.org/19/1/pdf_Formation_Mayaffre_Alexandrie_dec09_.pdf, site consulté le 21 septembre 2015).

données et sur les postulats épistémologiques qui sous-tendent l'usage de ces logiciels.

1.1. Les critères de sélection du corpus

Un corpus « n'a de sens, de valeur et de pertinence qu'au regard des questions qu'on va lui poser, des réponses que l'on cherche, des résultats que l'on va trouver[9] ». Nous souhaitons ici comprendre comment se caractérise le traitement médiatique de l'anorexie mentale dans la presse et quelles représentations les quotidiens construisent de cette pathologie. Pour cela, nous étudions les articles publiés dans *La Croix*, *Le Monde*, *Libération*, *L'Humanité* et *Le Figaro*, ces discours étant sélectionnés en référence à un « principe d'exemplarité » qui consiste à étudier des supports dont les positions éditoriales reposent sur des approches sociopolitiques différentes[10]. Notre corpus satisfait ainsi au critère de contrastivité. Il répond également à celui d'homogénéité puisque tous les articles retenus portent sur l'anorexie mentale à l'adolescence. Enfin, notre corpus respecte un critère de diachronicité. L'objectif étant d'étudier la « mise en visibilité » et « en lisibilité[11] » de l'anorexie dans la presse, il était nécessaire de disposer d'une période assez longue pour repérer les dynamiques de la construction médiatique du sujet. L'analyse porte donc sur une période de 14 années, du 1er janvier 1995 au 11 mars 2009. Les discours ont principalement été recueillis sur la base de données Lexis Nexis[12]. Une première lecture a permis d'exclure

9 Damon Mayaffre, « Les corpus *réflexifs* : entre architextualité et hypertextualité », *Corpus*, n° 1, 2002, http://corpus.revues.org/11, site consulté le 2 décembre 2014.

10 Isabelle Garcin-Marrou, *Des violences et des médias*, Paris, L'Harmattan, 2007, p. 16.

11 Bernard Delforce, « La responsabilité sociale du journaliste : donner du sens », *op. cit.*, p. 28.

12 Nous avons recueilli tous les articles contenant le mot anorexie dans le titre ou le texte. Pour la période de février 2008 à mars 2009, le corpus du *Monde* a été complété avec les archives du quotidien. Pour *Libération*, nous avons eu recours aux archives sur cédérom pour les années 1995-2000. Enfin, pour *L'Humanité*, nous avons consulté les archives sur le site internet entre 1995 et 2002.

ceux qui n'avaient pas de rapport avec la thématique choisie pour parvenir à un corpus de 131 articles (36 dans *La Croix*, 35 dans *Libération*, 27 dans *Le Figaro* et *Le Monde* et 6 dans *L'Humanité*). Ce corpus a d'abord été analysé à l'aide du logiciel Modalisa.

1.2. Modalisa : un logiciel de traitement d'enquêtes et de questionnaires

Modalisa[13] est un logiciel de traitement d'enquêtes et de questionnaires, créé en 1987, principalement utilisé pour des études de marketing, en sociologie ou encore en sciences de l'information et de la communication[14]. Dans le cadre de notre doctorat, nous avons eu recours à ce logiciel pour créer une enquête visant à caractériser la couverture médiatique des troubles de l'adolescence dans la presse quotidienne nationale. Notre démarche relève de l'analyse de contenu, définie par Berelson comme « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste des communications, ayant pour but de les interpréter[15] ». Le chercheur identifie un ensemble de catégories dont il repère les occurrences au fil de la lecture de son corpus, ces catégories sont ensuite comptabilisées pour donner lieu à des calculs statistiques. La difficulté réside

13 http://www.modalisa.com/index.php, site consulté le 10 octobre 2015.

14 Outre notre travail de doctorat, nous pensons également aux recherches d'Émilie Roche : « Étude des discours de presse écrite française sur la violence et la torture pendant la guerre d'Algérie : *Le Monde, L'Humanité, Le Figaro, L'Express, France-Observateur*, 1954-1962 », thèse de doctorat en sciences de l'information et de la communication, Lyon, Université Lyon 2, 2007, http://theses.univ-lyon2.fr/documents/lyon2/2007/roche_e#p=0&a=top, site consulté le 21 septembre 2015; de Lise Jacquez : « La controverse autour des expulsions des sans-papiers dans la presse française : une analyse des discours et des enjeux sociopolitiques », thèse de doctorat en sciences de l'information et de la communication, Lyon, Université Lyon 2, 2014, http://theses.univ-lyon2.fr/documents/lyon2/2014/jacquez_l/pdfAmont/jacquez_l_these.pdf, site consulté le 21 septembre 2015; et d'Aurélie Aubert : « Analyse de contenu et statistiques : une étude de la prise de parole des téléspectateurs au travers de leurs courriels », dans Camille Laville, Laurence Leveneur et Aude Rouger (dir.), *Construire son parcours de thèse – Manuel réflexif et pratique*, Paris, L'Harmattan, 2008, p. 97-104.

15 Laurence Bardin, *L'analyse de contenu*, Paris, Presses Universitaires de France, 1985, p. 40.

dans leur choix. En effet, la réussite de l'analyse de contenu « suppose que le système des catégories définies *a priori* est à la fois cohérent et pertinent[16] ». Pour cela, nous avons lu intégralement plusieurs fois notre corpus avant d'établir un questionnaire composé de 19 items. La première partie de l'enquête est relative au dispositif de l'article (titre de l'article, date de parution, auteur, etc.) tandis que la seconde porte sur le contenu des discours (trouble(s) de l'adolescence évoqué(s), acteurs, action(s), facteur(s) déclencheur(s) de la maladie, etc.). Ces catégories ont été définies en fonction de nos objectifs de recherche, de notre problématique et de notre méthodologie[17]. Notre grille d'enquête est la suivante :

16 Ludovic Lebart et André Salem, *Statistique textuelle*, Paris, Dunod, 1994, p. 14.

17 Les termes de facteurs déclencheurs et d'actions renvoient aux phases du schéma narratif greimassien (phases de manipulation et de performance). La notion d'acteur correspond à celle d'actant-sujet en sémiotique narrative (Algirdas Julien Greimas, *Du sens II : Essais sémiotiques*, Paris, Seuil, 1983).

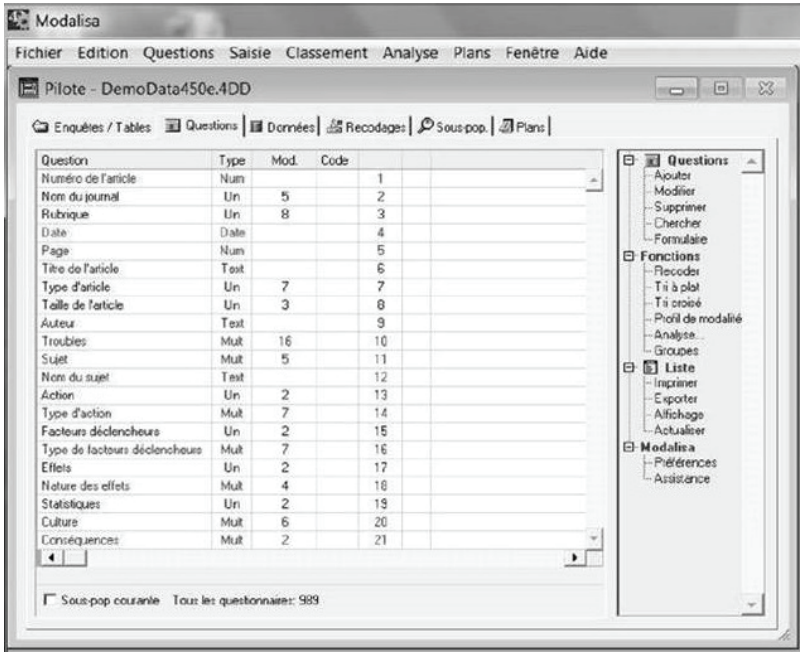
Grille d'enquête Modalisa

	Catégories	Type de données	Possibilités
1	Nom du journal	Numérique et à réponse unique	Le Monde, La Croix, Le Figaro, L'Humanité, Libération
2	Rubrique	Numérique et à réponse unique	France, Politique, Société, Médecine, Culture, Autres
3	Date	Textuelle	
4	Page	Numérique	
5	Titre de l'article	Textuelle	
6	Type de l'article	Numérique et à réponse unique	Brève, Article, Courrier, Interview, Editorial, Tribune, Autres
7	Auteur	Texte	
8	Troubles	Numérique et à réponses multiples	Anorexie, Boulimie, Alcool, Drogue, Toxicomanie, Suicide, Dépression, Fugue, Angoisse, Mutisme, Scarification/Automutilation, Repli sur soi, Mal-être, Phobie, Troubles, Autres
9	Sujet(s)	Numérique et à réponses multiples	Pouvoirs publics, Corps médical, Association, Adolescent, Autres
10	Nom du sujet	Textuelle	
11	Performance(s)	Numérique et à réponse unique	Oui/non
12	Type de performance	Numérique et à réponses multiples	Législation, Prévention, Prise en charge, Politiques publiques, Performance de l'adolescent, Autres
13	Facteurs déclencheurs	Numérique et à réponse unique	Oui/non
14	Type de facteurs déclencheurs	Numérique et à réponses multiples	Facteur génétique, Facteur familial, Facteur économique, Facteur socio-culturel, Facteur psychique, L'adolescence, Autres
15	Effets	Numérique et à réponse unique	Oui/non
16	Nature des effets (victime)	Numérique et à réponses multiples	Sur l'adolescent, Sur la famille, Sur la société, Autres
17	Statistiques	Numérique et à réponse unique	Oui/non
18	Culture	Numérique et à réponse unique	Film, Parution d'un ouvrage scientifique, Parution d'un roman, Émission télé ou radio, Conférence, Autres
19	Conséquences	Numérique et à réponses multiples	Justice/police

Ces questions sont intégrées dans l'enquête que crée le chercheur avec le logiciel. Elles se présentent de la façon suivante :

Figure 1

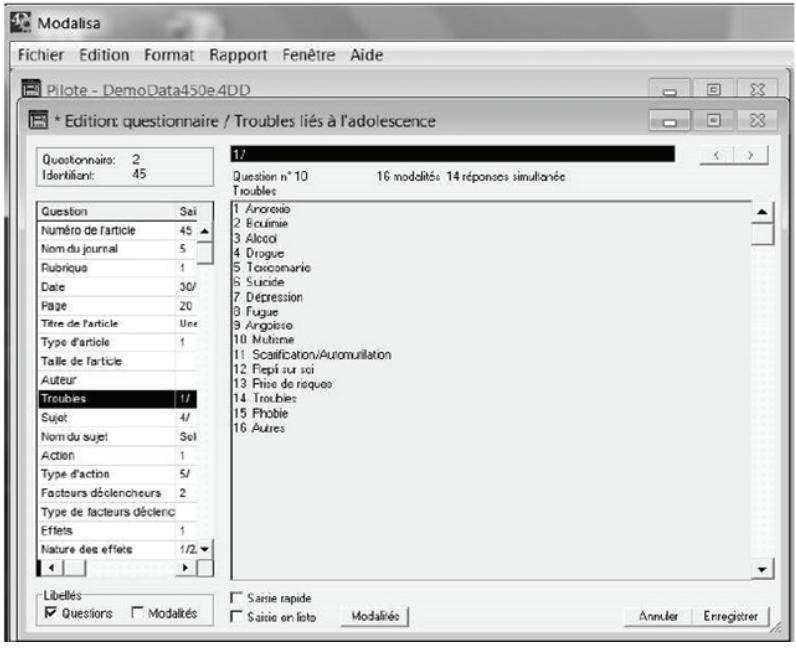
Questionnaire de l'enquête



Il existe plusieurs types de questions : numérique (la « réponse » est un chiffre, comme la page de publication d'un article), unique (le chercheur ne peut coder qu'une seule réponse parmi les différentes possibilités de réponse), multiple (plusieurs réponses sont possibles) ou textuelle (le texte de la réponse est tapé, comme le titre de l'article ou le nom de l'auteur). Ce questionnaire est rempli « manuellement » par le chercheur pour chaque article du corpus. Le processus de codage se présente de la façon suivante :

Figure 2

Écran de saisie d'un article[18]



Modalisa peut être considéré comme un logiciel « d'analyse manuelle » au sens où il « aide [...] au codage des unités de sens, facilite [...] la classification des données et fourni[t] une assistance précieuse lors de l'analyse et de la gestion des rapports[19] ». Christophe Lejeune propose le terme de « logiciel réflexif[20] » car la place accordée à la réflexion est primordiale. En effet, l'enquête est créée en fonction des objectifs de la recherche mais le travail du chercheur ne se limite pas à cette étape. Il doit également choisir l'unité de sens (ou « unité de décompte[21] ») pour le codage

18 Sur l'écran, figure la 8e question de l'enquête : quel(s) trouble(s) de l'adolescence évoque l'article? Les 16 modalités proposées figurent en colonne, à chacune étant associé un chiffre qui permet le codage de la réponse.

19 Normand Roy et Roseline Garon, *op. cit.*, p. 156.

20 Christophe Lejeune, « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches qualitatives*, n° 9, 2010, p. 26.

21 Ludovic Lebart et André Salem, *op. cit.*, p. 14.

des questions (s'agit-t-il du mot, de la phrase, du paragraphe, d'un élément syntaxique ou encore d'un élément sémantique[22]?). Il peut être amené à modifier l'enquête (et par conséquent, à revenir sur les questionnaires déjà saisis) si des éléments nouveaux et pertinents apparaissent à la lecture du corpus, au cours de la phase de codage. Une fois l'ensemble des articles codés, il est possible de générer différents types de graphiques (en tri à plat[23] ou croisé[24]) et de tableaux synthétisant les résultats. Les réponses aux questions de l'enquête peuvent toutes faire l'objet d'un traitement statistique à l'exception des réponses de type texte[25]. Il est également possible de travailler sur la totalité du corpus ou sur des sous-populations (dans notre cas, le corpus d'un journal par exemple). Le choix des représentations graphiques, des catégorisations, etc. dépend des hypothèses initiales du chercheur.

L'avantage majeur de Modalisa réside dans sa malléabilité en fonction des objectifs de recherche. Chaque axe d'analyse ou élément du discours peut faire l'objet d'une question dans l'enquête (thématiques principales, acteurs, arguments, causes d'un événement, experts sollicités, discours rapportés, désignation lexicale, etc.). Ainsi, dans le cas de notre travail, d'autres questions étaient envisageables : une étude argumentative des discours portant sur la prévention de l'anorexie, une analyse de la dénomination de l'anorexie comme problème de santé publique, l'étude des domaines scéniques[26] ou encore la place des experts scientifiques dans la médiatisation du sujet. Modalisa rend donc possible des parcours de lecture du corpus extrêmement riches et

22 *Ibid.*, p. 14.

23 Le tri à plat permet d'observer la distribution des différentes réponses à une question unique (par exemple, le nombre d'articles publiés pour chaque trouble de l'adolescence).

24 Le tri croisé permet de mettre en relation les résultats obtenus à deux questions de l'enquête.

25 Elles peuvent néanmoins faire l'objet d'un recodage qui consiste à regrouper les réponses pour les réduire à un nombre de modalités limité. La question initiale (de type texte) est transformée en question à réponse unique.

26 Cf. Patrick Charaudeau (dir.), *La médiatisation de la science. Clônage*, *OGM, manipulations génétiques*, Bruxelles, Éditions de Boeck, 2008, p. 26. La notion de domaine scénique renvoie au cadre de questionnement d'un problème et aux acteurs désignés dans les discours pour le prendre en charge.

variés. Cependant, quels que soient les axes d'analyse choisis, l'usage du logiciel va de pair avec une réflexion approfondie sur les questions et les modalités composant l'enquête. Cette étape est particulièrement longue et nécessite une bonne connaissance du corpus et un cadre d'analyse (problématique, hypothèses, voire concepts théoriques) déjà bien défini. Autrement dit, ce logiciel est particulièrement adapté pour les démarches hypothético-déductives mais moins pour des analyses inductives. Nous allons maintenant voir en quoi un logiciel d'« analyse automatique » de discours comme Iramuteq se distingue d'un logiciel d'« analyse manuelle ».

1.3. Iramuteq : un logiciel d'« analyse automatique » de discours

Iramuteq est un logiciel d'analyse de textes et de questionnaires, développé depuis 2008, dont l'usage est encore confidentiel. Toutefois, des travaux en sciences de l'information et de la communication, sociologie, psychologie ou encore médecine en font le cœur de leur cadre méthodologique. Ce logiciel s'inscrit dans la catégorie des logiciels « automatiques » : il « fait la plus grande partie de l'analyse avec un minimum d'interventions de la part du chercheur » et « permet surtout de faire de l'analyse lexicométrique avec des statistiques textuelles[27] ». Autrement dit, et contrairement à Modalisa, l'analyse produite ne résulte pas d'une catégorisation *a priori* opérée par le chercheur, en fonction de ses objectifs de recherche, mais d'un ensemble de procédés statistiques.

Le logiciel s'appuie sur une fragmentation du corpus qui est découpé en textes[28], eux-mêmes divisés en segments au sein desquels sont repérées les formes co-occurrentes[29]. La corrélation

27 Normand Roy et Roseline Garon, *op. cit.*, p. 155.

28 Dans notre corpus, chaque article représente un texte mais d'autres paramétrages sont possibles.

29 La co-occurrence est une « association statistiquement significative de deux items (en général deux mots) dans une fenêtre déterminée du texte (en général le paragraphe) » (Damon Mayaffre, « Plaidoyer en faveur de l'Analyse de Données co(n)Textuelles. Parcours cooccurrentiels dans le discours présiden-

récurrente entre mots co-occurrents permet d'obtenir un aperçu des thématiques présentes dans les discours. En effet, c'est « la co-occurrence des mots pleins appartenant à un même segment [qui] est une trace possible de l'acte d'énonciation, et donc de son contenu[30] ». Par un procédé de classification descendante hiérarchique[31], ces segments sont ensuite regroupés en fonction de la distribution différenciée de leur vocabulaire. Le chercheur obtient alors un nombre x de classes (ou « mondes lexicaux[32] ») qui sont autant de « points de vue » sur l'objet étudié. À chacun de ces univers thématiques[33] est associée une liste de formes (ou mots)[34], dont le degré d'appartenance à la classe est indiqué par

---

tiel français (1958-2014) », 12e Journées internationales d'Analyse statistique des Données Textuelles, 2014, p. 18, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/01-JADT2014.pdf, site consulté le 21 septembre 2015.

30 Max Reinert, « Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours », *Langage et société*, n° 121-122, 2007, p. 193.

31 La technique de la classification hiérarchique descendante est détaillée dans Max Reinert, « La méthode d'analyse exploratoire des données textuelles "Alceste" et le problème de l'analyse de contenu », *Les Cahiers de Jérico-st*, n° 4, Presses de l'Université de Tours, 2004, p. 79-82.

32 Max Reinert définit la notion de « monde lexical » de la façon suivante : « Un énoncé traduit donc davantage un *point de vue* particulier […]. Notre hypothèse principale consiste justement à considérer le vocabulaire d'un énoncé particulier comme une trace pertinente de ce *point de vue.* […] Nous appelons *mondes lexicaux*, les traces les plus prégnantes de ces activités dans le lexique » (Max Reinert, « Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars », *Langage et société*, n° 66, 1993, p. 11).

33 La notion d'univers thématique désigne la thématique à laquelle se rapportent la ou les classes issues de la classification hiérarchique descendante. Nous l'employons comme synonyme de monde lexical ou univers lexical. D'autres dénominations existent : « univers sémantique » ou « contexte sémantique » (Olivier Laügt, « Le SRAS dans *Le Monde* : un agent double? », *Les Cahiers du journalisme*, n° 15, 2006, p. 89), « classe sémantico-thématique » ou « classe thématico-sémantique » (Claire Blandin *et al.*, « Femmes et alcool dans la presse écrite française (1980-2012) : construction sociale des problèmes de santé publique et des rapports de genres », *Les cahiers de l'IREB*, n° 21, 2013, p. 154, http://www.ireb.com/sites/default/files/Cahiers%2021.pdf, site consulté le 17 septembre 2015).

34 En statistique textuelle, une « suite de caractères non-délimiteurs bornée à ses deux extrémités par des caractères délimiteurs est une occurrence. Deux suites identiques de caractères non-délimiteurs constituent deux occurrences d'une même forme » (Ludovic Lebart et André Salem, *op. cit.*, p. 36).

le chi 2 (cf. figure 3 ci-dessous). Précisons que, pour cette analyse, seuls les mots pleins sont retenus (noms, adjectifs, verbes, adverbes), les autres étant considérés comme des mots-outils. Ces mots pleins peuvent être lemmatisés[35] autrement dit, les différentes formes d'un mot sont ramenées à une forme unique, celle du lemme[36]. Notre analyse se fonde sur cette classification du corpus en mondes lexicaux mais d'autres fonctionnalités sont proposées par Iramuteq[37].

Figure 3

Profil d'une classe issue de la classification hiérarchique descendante[38]

Description Corpus Anorexie Alceste - revue Iramuteq2_corpus_7 | Classification - Corpus Anorexie Alceste

CHD | Profils | AFC

classe 1 (107/483 - 22.15%) | classe 2 (83/483 - 17.18%) | classe 3 (92/483 - 19.05%) | classe 4 (72/4

n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme
0	16	23	69.57	31.48	nom	mère
1	15	21	71.43	30.91	nr	marcel_rufo
2	26	52	50.0	26.2	ver	aller
3	16	28	57.14	21.1	nom	parent
4	7	8	87.5	20.14	nom	rendez_vous
5	8	11	72.73	16.69	nom	temps
6	4	4	100.0	14.17	nom	maison
7	4	4	100.0	14.17	nom	bureau
8	4	4	100.0	14.17	nr	nantes
9	5	6	83.33	13.19	ver	aimer
10	5	6	83.33	13.19	ver	retrouver
11	5	6	83.33	13.19	nom	parole
12	5	6	83.33	13.19	nom	juin
13	6	8	75.0	13.17	ver	consulter
14	11	20	55.0	13.05	nom	médecin
15	8	13	61.54	12.02	nom	jour
16	3	3	100.0	10.61	ver	raconter
17	3	3	100.0	10.61	nom	pédiatre
18	3	3	100.0	10.61	ver	culpabiliser
19	3	3	100.0	10.61	nom	envie

35 Le choix est donné au chercheur dans Iramuteq.

36 Les verbes conjugués sont ramenés à leur infinitif, les substantifs au singulier, les adjectifs au masculin singulier et les formes élidées à la forme sans élision.

37 Le logiciel permet d'obtenir un index des formes composant le corpus, de procéder à une analyse factorielle des correspondances, de réaliser des nuages de mots ou encore une analyse des spécificités et des similitudes.

38 Les formes (à droite) sont classées en fonction de leur chi 2 : plus celui-ci est élevé, plus le mot est représentatif de la classe.

Au vu du fonctionnement d'un logiciel d'« analyse automatique », nous pouvons nous interroger sur la place du chercheur dans ce type d'analyse. *A priori*, elle semble minime. En effet, alors qu'il interprète déjà son corpus en créant une grille d'enquête avec Modalisa, ici ce sont des outils statistiques qui comptabilisent les formes et « apprécient » leur distribution dans les textes, sans prendre en compte leur sens[39]. Les résultats doivent donc être utilisés avec précaution et leur interprétation nécessite un retour systématique au texte – permis par le concordancier[40] – ainsi qu'une bonne connaissance du corpus. Si la réflexion du chercheur est donc indispensable en aval, elle existe toutefois aussi en amont. En effet, le recours à la statistique textuelle implique une préparation et une codification du corpus avant son importation dans le logiciel. Lors de la phase de préparation, le chercheur peut être amené à modifier certains éléments des discours pour que l'ensemble des formes graphiques soient reconnues par le logiciel. Ainsi, et comme le préconisent les concepteurs d'Iramuteq, il est indispensable de corriger les fautes d'orthographe et d'harmoniser les mots écrits différemment[41]. Il

39 Par exemple, le logiciel ne fait pas de distinction entre les homographes. Dans notre corpus, le mot « mode » renvoie à la fois à l'univers du mannequinat mais est aussi un synonyme de « façon » ou « manière » (« le mode d'expression »). Pour le logiciel, ces occurrences sont comptabilisées ensemble. Toujours par rapport au sens, le logiciel ne peut pas non plus détecter l'ironie ou le second degré.

40 Le concordancier est une fonction du logiciel qui permet d'obtenir la liste de toutes les occurrences d'un mot en contexte (Bénédicte Pincemin *et al.*, « Concordanciers : thème et variations », 8e Journées internationales d'Analyse statistique des Données Textuelles, 2006, https://halshs.archives-ouvertes.fr/halshs-00154100/file/pincemin_al_jadt06_texte.pdf, site consulté le 21 septembre 2015).

41 En effet, Pierre Ratinaud et Pascal Marchand stipulent qu'« une intervention morphologique s'impose donc pour permettre la reconnaissance des formes [...]. Les formes mal orthographiées ne peuvent pas être lemmatisées » (Pierre Ratinaud et Pascal Marchand, « Recherche improbable d'une homogène diversité : le débat sur l'identité nationale », *Langages*, n° 187, 2012, p. 96-97). De même, dans une recherche collective, ils expliquent que la variabilité dans la graphie de leur corpus est telle « qu'une analyse automatique serait considérée comme peu stable : erreurs orthographiques, approximations syntaxiques, termes orduriers, etc. Une première étape de travail devait alors en rétablir l'homogénéité » (Alexia Ducos *et al.*, « Classification

est également possible de regrouper plusieurs termes en une seule expression[42]. Ces « opérations d'homogénéisation[43] » ne sont pas toujours aisées et reposent sur des choix intellectuels qui conditionnent les résultats finaux. Elles nécessitent donc une (voire plusieurs) lecture(s) attentive(s) du corpus. Enfin, la codification consiste à choisir les variables qui seront associées aux discours et introduites sous la forme d'une ligne étoilée[44]. Elles dépendent des objectifs de recherche et doivent être définies avant le traitement du corpus.

Avec un logiciel d'« analyse automatique », le texte est considéré comme un « sac de mots[45] » au sens où le découpage qui en est fait équivaut à une lecture fragmentée et non linéaire. La classification hiérarchique descendante est indépendante de la succession des articles dans le corpus, de l'ordre des segments et de celui des mots. Cette « lecture » a pour principal avantage de proposer un autre regard sur le corpus et de faire surgir des éléments que le chercheur n'aurait peut-être pas perçus avec une simple lecture flottante. Cette fragmentation des discours pose cependant la question de la signification : alors qu'avec Modalisa, chaque indicateur est signifiant (par rapport à une problématique, des hypothèses, etc.), que mesurent ces procédés statistiques? La lexicométrie calcule des fréquences de formes, repère des co-occurrences et dégage des spécificités. Christophe Lejeune rappelle qu'« une

---

d'un corpus hétérogène : la page *Facebook* de soutien au "bijoutier de Nice" (septembre 2013) », 12e Journées internationales d'Analyse statistique des Données Textuelles, 2014, p. 227, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/18-JADT2014.pdf, site consulté le 21 septembre 2015. Par exemple, dans notre corpus, la campagne publicitaire d'Oliviero Toscani sur l'anorexie est appelée No-l-ita ou Nolita selon les articles. Nous avons donc harmonisé les différentes graphies de ce terme.

42 Par exemple, dans l'expression « ministère de la santé », les différents mots peuvent être regroupés par le signe "_" pour indiquer au logiciel qu'ils ne doivent pas être séparés. Ils sont alors traités (et comptés) comme une forme unique.

43 Pierre Ratinaud et Pascal Marchand, *op. cit.*, p. 94.

44 Par exemple, dans une analyse de discours de presse, les variables peuvent être le journal et la date de parution de l'article. Pour chaque discours, une ligne étoilée introduira ces paramètres : **** *journal_LC *date_2005

45 Ludovic Lebart et André Salem, *op. cit.*, p. 146.

telle approche quantitative repose sur l'hypothèse que les phénomènes de récurrence ont une pertinence »[46]. De même, la sociologue Monique Dalud-Vincent souligne que « c'est la régularité et non la singularité qui prime »[47]. Ainsi, une forme ayant un faible nombre d'occurrences n'est pas retenue dans la classification hiérarchique descendante[48]. En outre, la fragmentation du texte – « opération fondamentale de la méthode[49] » – est arbitraire, ce qui pose question puisqu'un chercheur averti peut modifier le paramétrage par défaut, choisissant ainsi entre plusieurs découpages possibles. Pour Valérie Delavigne, « la construction de classes peut laisser croire que le logiciel livre une "vérité intrinsèque" sur le corpus, mais il s'avère que, dès lors que l'on change quelques paramètres (modification des variables par exemple), ces classes peuvent changer[50] ». Dans Iramuteq, la classification hiérarchique descendante peut être simple sur segment de texte[51], simple sur texte[52] ou double sur regroupements de segments de texte[53]. Nous reviendrons sur les conséquences de ces choix techniques dans la partie suivante. Contrairement à Modalisa, ce logiciel semble *a priori* plus adapté pour des approches inductives. L'analyse de notre corpus de presse va nous permettre de préciser cette intuition.

46 Christophe Lejeune, *op. cit.*, p. 21.

47 Monique Dalud-Vincent, « Alceste comme outil de traitement d'entretiens semi-directifs : essai et critiques pour un usage en sociologie », *Langage et société*, n° 135, 2011, p. 11.

48 Dans Iramuteq, une forme est prise en compte à partir de trois occurrences, ce seuil pouvant être modifié.

49 Max Reinert, « La tresse du sens et la méthode "Alceste". Application aux "Rêveries du promeneur solitaire" », 5e Journées internationales d'Analyse Statistique des Données Textuelles, 2000, p. 3, http://lexicometrica.univ-paris3.fr/jadt/jadt2000/pdf/31/31.pdf, site consulté le 21 septembre 2015.

50 Valérie Delavigne, « Alceste, un logiciel d'analyse textuelle », Texto! Textes et cultures, *Équipe Sémantique des textes*, 2003, p. 5, https://hal.archives-ouvertes.fr/hal-00924168/document, site consulté le 21 septembre 2015.

51 La classification porte sur les segments de texte qui peuvent être définis par un nombre x d'occurrences, de caractères ou de paragraphes. Le paramétrage par défaut est fixé à un seuil de 40 occurrences.

52 Les textes sont considérés dans leur intégralité, le logiciel rapproche ceux qui sont proches d'un point de vue lexical.

53 Le découpage s'opère sur des regroupements de segments de texte. L'opération est réalisée deux fois pour garantir la stabilité des résultats en variant le nombre de formes actives dans les regroupements de segments de texte.

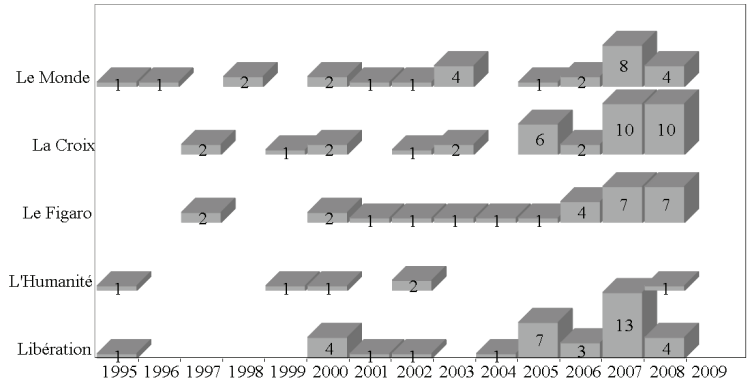
2. Les représentations médiatiques de l'anorexie mentale dans la presse quotidienne nationale

2.1. L'analyse avec Modalisa : une caractérisation de la couverture médiatique du sujet

Nous avons eu recours au logiciel Modalisa pour caractériser le traitement médiatique de l'anorexie mentale (phases de médiatisation, pic(s) médiatique(s) commun(s) aux journaux, etc.) puis pour étudier le rubriquage et la configuration du sujet comme problème public[54]. Concernant la façon dont les quotidiens mettent « en visibilité[55] » ce sujet, le tri croisé permet d'obtenir les résultats suivants :

Figure 4

Répartition du volume d'articles sur l'anorexie mentale par journal entre 1995 et 2009



54 Défini par Erik Neveu comme la « transformation d'un "fait social" quelconque en enjeu de débat public et/ou d'intervention étatique. Du plus tragique au plus anecdotique, tout fait social peut potentiellement devenir un "problème social" s'il est constitué par l'action volontariste de divers opérateurs (presse, mouvements sociaux, partis politiques, lobbies, intellectuels...) comme une situation problématique devant être mise en débat et recevoir des réponses en termes d'action publique (budgets, règlementation, répression...) » (Erik Neveu, « L'approche constructiviste des "problèmes publics" – Un aperçu des travaux anglo-saxons », *Études de communication*, n° 22, 1999, p. 42).

55 Bernard Delforce, *Les Cahiers du journalisme*, *op. cit.*, p. 28.

La médiatisation du sujet est irrégulière et hétérogène : selon les journaux, entre 0 et 13 articles sont publiés annuellement avec un pic médiatique en 2007-2008. Plusieurs événements expliquent cette entrée « soudaine » de l'anorexie sur la scène médiatique. L'été 2006, deux mannequins anorexiques meurent. Ces décès donnent lieu à une mobilisation des professionnels du milieu de la mode pour réglementer les défilés de mode. En 2007, la campagne publicitaire Nolita d'Oliviero Toscani suscite la polémique : une jeune française anorexique, au corps décharné, pose nue pour alerter sur les dangers de la maladie. En avril 2008, le gouvernement français propose une charte d'engagement sur l'image corporelle incitant à valoriser la diversité corporelle. La signature de ce document est suivie d'une proposition de loi visant à réprimer l'incitation à l'anorexie ou à la recherche d'une maigreur extrême, qui concerne notamment les blogs pro-anorexiques[56]. Ces différents événements sont relayés par la presse. L'intérêt soudain pour l'anorexie peut s'expliquer par un phénomène de contagion-assimilation tel que le définissent Jacques Noyer et Bernard Delforce :

> à partir d'un enjeu social déterminé et par un fonctionnement de proche en proche, des discours sociaux relatifs à une catégorie particulière d'occurrence-événement [...] peuvent tisser des liens avec d'autres types d'occurrences-événements [...]. Ce phénomène de "contagion/assimilation" peut faire émerger des discours sociaux relatifs, cette fois, à plusieurs types d'occurrences-événements sur lesquels ces discours prennent appui, en même temps qu'ils en créent une perception globalisante, c'est-à-dire comme occurrences-événements finalement d'un même type[57].

La polémique déclenchée par la mort de deux mannequins a donné aux journaux un cadre à partir duquel appréhender l'anorexie, les années 2007-2008 constituant un moment où se matérialise le croisement de différents enjeux. L'étude de la place accordée au facteur déclencheur socioculturel de cette maladie,

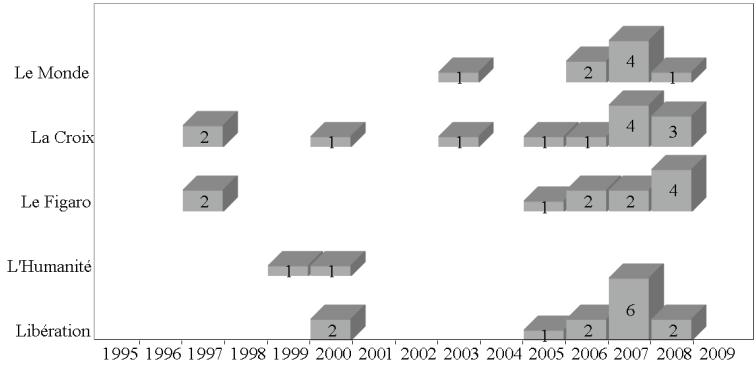
56 Adoptée en première lecture par l'Assemblée nationale, le texte n'a toujours pas été promulgué.

57 Bernard Delforce et Jacques Noyer, « Pour une approche interdisciplinaire des phénomènes de médiatisation : constructivisme et discursivité sociale », *Études de communication*, n° 22, 1999, p. 23.

dans les discours, le confirme. Le graphique suivant indique le nombre d'articles évoquant ce facteur déclencheur en fonction des années (en abscisse) et des quotidiens (en ordonnée)[58].

Figure 5

Évolution de la place accordée au facteur déclencheur socioculturel de l'anorexie dans les discours



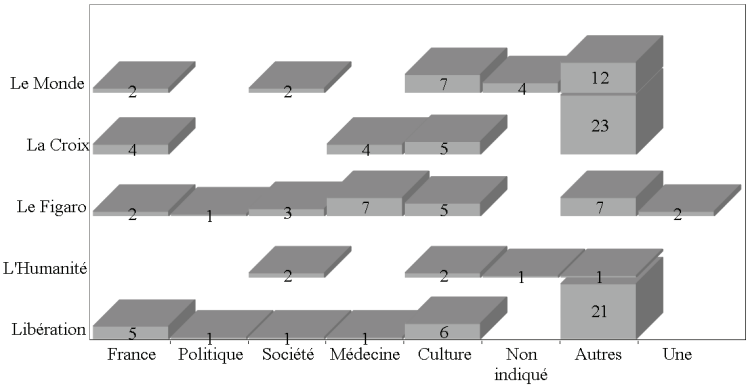
La référence à une origine socioculturelle de l'anorexie mentale coïncide avec le pic médiatique précédemment identifié et confirme ce « phénomène de contagion/assimilation ». Nous pouvons nous demander si cette émergence de l'anorexie dans la presse quotidienne s'accompagne d'un cadrage en termes de problème de santé publique. L'étude de l'évolution du rubriquage[59] constitue une façon de répondre à cette question. Le graphique suivant indique le nombre d'articles publiés dans chaque rubrique (en abscisse) selon les quotidiens (en ordonnée).

58 Le graphique est réalisé à partir de la question 14 de la grille d'enquête « Types de facteurs déclencheurs ».

59 Les différentes rubriques codées dans l'enquête sont les suivantes : France, Politique, Société, Médecine, Culture, Autres et Non indiqué. La modalité « Autres » regroupe les rubriques qui ne sont pas communes à l'ensemble des titres de presse.

Figure 6

Rubriquage des articles sur l'anorexie dans la presse quotidienne



Contrairement à d'autres problèmes de santé publique, l'anorexie mentale ne fait pas l'objet d'une « migration[60] » des pages scientifiques vers les pages plus prestigieuses des quotidiens (les rubriques France et Politique sont peu investies par les journaux). Le sujet ne donne pas non plus lieu à une « extension rubriquale[61] » mais reste dispersé entre les différentes rubriques des journaux. L'absence d'une évolution claire du cadrage informatif et la diversité constante du rubriquage tout au long de la période souligne la difficulté pour la presse à intégrer l'anorexie mentale dans des cadres de référence préétablis[62]. Ce sujet ne semble donc *a priori* pas construit comme un problème public par la presse.

60 Sophie Moirand, « Du traitement différent de l'intertexte selon les genres convoqués dans les événements scientifiques à caractère politique », *Semen*, n° 13, 2001, p. 101-102.

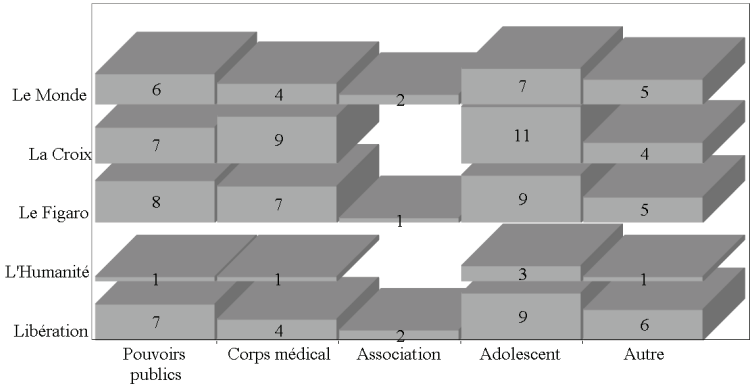
61 Jacques Noyer, « La couverture du sida dans la presse française de 1982 à 1989 à travers trois quotidiens nationaux (*Le Figaro*, *Libération*, *Le Monde*) : approches de la notion d'événement », thèse de doctorat en sciences de l'information et de la communication, Lille, Université Lille 3, 1994, p. 116.

62 Dans un quotidien, « le système des rubriques [...] est un cadre de références qui constitue une lecture de notre environnement social : il en classe les

L'étude de la place accordée à la figure des pouvoirs publics dans les discours le confirme. En effet, dans leur analyse de la couverture médiatique du cancer dans la presse d'information générale, Leila Azeddine, Gersende Blanchard et Cécile Poncin rappellent que la médiatisation d'un problème de santé publique se caractérise par une augmentation du nombre d'articles consacrés au sujet, « une présence plus marquée de la sphère politique dans le contenu des articles » et une évolution du rubriquage[63]. Dans notre corpus sur l'anorexie, la présence de la figure des pouvoirs publics est relativement faible[64] comme le montre le graphique suivant :

Figure 7

Place de la figure des pouvoirs publics dans les discours sur l'anorexie mentale



L'anorexie mentale n'est pas un sujet très médiatisé même s'il retient l'attention de la presse quotidienne en 2007-2008. En outre, cette maladie n'est pas construite comme un problème de

composantes qui deviennent elles-mêmes des intruments de classement de l'actualité » (*Ibid.*, p. 114).

63 Leila Azeddine, Gersende Blanchard et Cécile Poncin, « Le cancer dans la presse d'information générale. Quelle place pour les malades? », *Questions de communication*, n° 11, 2007, p. 117-118.

64 Rappelons que notre corpus compte 131 articles.

santé publique par les journaux comme le montre l'analyse effectuée à partir de trois indicateurs[65] de notre grille d'enquête Modalisa, caractéristiques d'une configuration en termes de problème public. Partant de ces résultats, nous avons constitué un second corpus, plus restreint, sur lequel porte l'analyse avec le logiciel Iramuteq.

2.2. L'analyse avec Iramuteq : une mise en lumière des différents mondes lexicaux structurant les discours

Un recensement des articles ou ouvrages s'appuyant sur Iramuteq montre que, selon les choix opérés par les auteurs, l'objectif est d'identifier des univers thématiques, comprendre la configuration et la structuration d'une polémique[66] ou encore mettre en lumière des grandes tendances discursives sur un sujet pour ensuite travailler sur un corpus plus restreint[67]. Ici, notre objectif est de repérer les thématiques autour desquelles s'articulent les discours de presse et la façon dont elles évoluent. L'analyse avec Iramuteq confirme-t-elle le « phénomène de contagion/assimilation » identifié avec Modalisa? Existe-t-il une « temporalité thématique[68] » des récits médiatiques? Est-elle la même pour tous les journaux? Pour répondre à ces questions, nous avons introduit les variables « journal » (les différentes modalités correspondent aux cinq quotidiens composant le corpus) et « date » (les modalités sont alors l'année de publication des articles) lors de la codification des discours.

65 Rubrique, date et sujet(s).

66 Alexia Ducos *et al., op. cit.*

67 Annelise Touboul *et al.*, « La disparité des modes de traitement journalistiques et des énonciations éditoriales sur le Web. Le cas d'un sondage sur Marine Le Pen et la Présidentielle de 2012 », *Réseaux*, n° 176, 2012, p. 73-103; Lucie Loubère, « Le traitement des TIC dans les discours politiques et dans la presse », 12e Journées internationales d'Analyse statistique de Données Textuelles, 2014, p. 433-445, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/36-JADT2014.pdf, site consulté le 21 septembre 2015.

68 Laurence Joselin *et al.*, « Dynamiques temporelles de la pandémie de grippe A/H1N1 dans la presse écrite francophone », 12e Journées internationales d'Analyse statistique de Données Textuelles, 2014, p. 299-310, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/25-JADT2014.pdf, site consulté le 21 septembre 2015.

L'étape de préparation du corpus a été relativement longue. Par souci de cohérence, et pour ne pas biaiser les résultats, nous avons :

- regroupé les mots désignant des organismes, des institutions[69] et des personnes[70],
- regroupé les termes constituant une expression[71],
- harmonisé les mots qui renvoient à un référent identique mais de façon différente[72],
- réécrit les termes familiers ou abrégés dans leur forme intégrale[73],
- supprimé les éléments relatifs au dispositif de l'article[74],
- détaillé les sigles qui ne sont pas reconnus par le logiciel[75],
- écrit l'ensemble des chiffres en lettres pour qu'ils puissent être intégrés à l'analyse[76].

Ces choix ont nécessité plusieurs lectures et ébauches d'analyse pour aboutir au « corpus final ». En effet, chaque décision conditionne partiellement les résultats et les fait évoluer. Ainsi, nous avions choisi de regrouper les termes « anorexie » et « mentale » par souci de cohérence intellectuelle. Cependant, après une première analyse, nous nous sommes rendu compte que l'expression « anorexie_mentale » était dissociée du terme « anorexie » (employé seul) dans le comptage des fréquences,

69 ministère_de_la_santé, bureau_de_vérification_de_la_publicité, maison_des_adolescents, etc.

70 Philippe_Jeammet, Maurice_Corcos, Xavier_Bertrand, Valérie_Boyer, Isabelle_Caro, etc.

71 Notamment pro_anorexi(qu)e, pouvoirs_publics.

72 Par exemple, « Ana Carolina » et « Ana Carolina Reston » désignent la même personne et le nom doit donc être écrit de façon identique pour que le logiciel comptabilise ces occurrences ensemble.

73 « toxico(s) » devient « toxicomane(s) »; « ado(s) » => « adolescent(s) »; « pub » => « publicité », « cm » => « centimètres », etc.

74 Rubrique, numéro de page, auteur (ces éléments étant conservés dans un document annexe).

75 UMP => Union_pour_un_mouvement_populaire; IMC => Indice_de_masse_corporelle; TCA => Trouble du comportement alimentaire, etc.

76 52 kilos => cinquante_deux kilos; 30 000 euros => trente_mille euros, etc.

le logiciel les « comprenant » comme deux mots différents. Nous sommes donc revenue sur cette codification initiale. Cet exemple constitue un exemple parmi d'autres de l'ensemble des modifications, des réflexions et des choix intellectuels qui précèdent la phase d'analyse avec le logiciel.

Comme cela a été expliqué précédemment, Iramuteq propose différents outils d'analyse. Nous nous limitons à la classification hiérarchique descendante qui permet de mettre en lumière les différents mondes lexicaux structurant les discours. Nous comparons les résultats obtenus sur un même corpus[77] à partir de plusieurs paramétrages, afin de discuter l'impact des choix techniques (paramétrages) et intellectuels (interventions sur le corpus) sur ces résultats :

- la première classification a été obtenue en utilisant les paramètres par défaut (classification simple sur segments de texte de 40 occurrences avec un nombre de classes pour la phase terminale fixé à 10), après réalisation des « opérations d'homogénéisation » précédemment mentionnées;
- dans la deuxième, nous avons optimisé le pourcentage de segments de texte classés en diminuant le nombre de classes pour la phase terminale à 6;
- la troisième classification fait suite à deux modifications dans le corpus. La première concerne l'expression « pro-ana » – raccourci pour « pro-anorexique » – souvent employée par les quotidiens entre guillemets. Nous trouvons également l'expression « pro-anorexique(s) » mais de façon moins fréquente. Nous avons donc transformé « pro-ana » en « pro-anorexique(s) » afin d'homogénéiser les discours. La seconde modification concerne le terme « anorectique » qui apparaît à plusieurs reprises (et uniquement) dans un article du *Figaro*. Ce terme scientifique est synonyme d'anorexique (comme adjectif ou substantif). Son emploi s'explique par le fait que l'auteur de l'article (Monique Vigy) est médecin et s'intéresse à un colloque sur

77 Le second corpus compte 13 articles dans *La Croix*, 7 dans *Le Monde*, 7 dans *Le Figaro*, 9 dans *Libération* et 3 dans *L'Humanité*.

les causes des troubles du comportement alimentaire. Publié dans la rubrique « La vie scientifique », la tonalité du discours est donc différente du reste des articles. Nous avons remplacé ce terme par celui d'« anorexique », également pour homogénéiser les discours. Toutefois, ce choix peut être discuté en fonction des objectifs scientifiques poursuivis. Un chercheur travaillant sur le genre des articles ou sur le lien entre le statut des journalistes (médecin, expert, etc.) et le lexique employé dans les discours aurait intérêt à conserver le terme « anorectique ».

Nous synthétisons les profils des trois classifications hiérarchiques descendantes obtenues dans des tableaux distincts. La première colonne indique la classe[78] et la seconde précise son poids par rapport à l'ensemble des segments de texte du corpus classés. Viennent ensuite les mots les plus représentatifs de chacune de ces classes (troisième colonne) et les variables qui y sont associées (dernière colonne). Nous menons une analyse comparative et globale de ces tableaux afin de montrer les résultats qui sont communs aux différentes classifications malgré les variations techniques et intellectuelles. Nous pointons aussi des écarts qui nous amèneront à formuler des hypothèses quant au fonctionnement du logiciel.

Dans la première classification hiérarchique descendante, 490 segments de texte sur 635 sont classés, soit 77,17 % du corpus.

78 Le numéro donné par le logiciel est arbitraire et n'est pas relatif à l'importance quantitative de la classe dans le corpus.

Tableau 1

Résultats de la première classification hiérarchique descendante

Classes	Volume	Formes actives les plus représentatives[i]	Modalités associées[ii]
1 **C**	107 ST 21.84 % du corpus	Kilo (chi 2 : 59,56), poids, mannequin, taille, perdre, gros, top_modèle, époque, femme, long, vitrine, défiler, magazine, etc.	Date : 2006, 2000 Journal : *L'Humanité*
2 **B**	92 ST 18.78 %	Trouble (chi 2 : 97,8), alimentaire, mental, comportement, pathologie, pour_cent, servir, soigner, xavier_darcos, exister, physique, problème, maladie, élève, sévère, dépression, compliqué, moyen, entourage, etc.	Date : 2003, 2001 Journal : *Le Figaro*
3	103 ST 21.02 %	Patient (chi 2 : 33,11), familial, hospitalisation, famille, suivre, nourriture, repas, façon, table, séparation, tendance, difficile, milieu, rencontrer, revanche, psychologique, exprimer, discussion, etc.	Date : 1997, 2003, 2005 Journal : *Le Figaro*
4 **D**	119 ST 24.29 %	Moment (chi 2 : 27,16), marcel_rufo, juin, sortir, arriver, arrêter, valoir, partir, etc.	Date : 2007, 1999 Journal : *Libération*
5 (la plus différenciée) **A**	69 ST 14.08 %	Publicité (chi 2 : 139,4), charte, marque, oliviero_toscani, média, ministre_de_la_santé, campagne, image, vêtement, professionnel, signer, roselyne_bachelot, nu, no_li_ta, décharner, diversité, engagement, mode, jean_pierre_poulain, volontaire, photographe, italien, corps, corporel, égide, réaliser, parrainer, benetton, ministère_de_la_santé, conduite, valorisation, sociologue, promouvoir, privilégier, positif, choquer, organisation, choquer, italie, esthétique, anorexie, mesure, groupe, etc.	Date : 2007, 2008 Journal : *Le Monde*

[i] Nous n'avons pas retranscrit intégralement la liste de toutes les formes associées aux différentes classes. Les formes qui figurent dans le tableau sont les plus représentatives; autrement dit, celles dont le degré de signification du chi 2 est inférieur à 0,0001.

[ii] Les modalités des variables associées aux classes sont données dans l'ordre du logiciel, c'est-à-dire en fonction de leur chi 2. Par exemple, pour la classe 1, la modalité « 2000 » a un chi 2 inférieur à la modalité « 2006 », ce qui signifie que ce monde lexical est moins représentatif de l'année 2000 que de l'année 2006.

Dans la deuxième classification hiérarchique descendante, 506 segments de texte sur 635 sont classés, soit 79,69 % du corpus.

Tableau 2

Résultats de la deuxième classification hiérarchique descendante

Classes	Volume	Formes actives les plus représentatives	Variables associées
1 **C**	140 ST 27.67 % du corpus	Mannequin (chi 2 : 62,81), kilo, taille, perdre, poids, gros, magazine, époque, top_modèle, etc.	Date : 2000, 2006, 2007, 1999 Journal : *L'Humanité*
2 **D**	119 ST 23.52 %	Moment (chi 2 : 28,39), marcel_rufo, isabelle_caro, juin, sortir, arrêter, heure, coup, réagir, partir, jouer, retrouver, etc.	Date : 2007, 1999 Journal : *Libération*
3 (la plus différenciée) **B**	247 ST 48.81 %	Trouble, alimentaire, mental, comportement, patient, maladie, etc.	Date : 2003, 1997, 2005 Journal : *Le Figaro*

Les profils de classe suivants ont été obtenus après avoir intégré dans le corpus les modifications sémantiques précédemment mentionnées. Nous obtenons 488 segments de texte classés sur 635, soit 76,85 % du corpus.

Tableau 3

Résultats de la troisième classification hiérarchique descendante

Classes	Volume	Formes actives les plus représentatives	Variables associées
1 **D**	130 ST 28.64 % du corpus	Moment (chi 2 : 28,11), marcel_rufo, parler, sortir, rendez_vous, consulter, arrêter, retrouver, aller, sentir, occuper, etc.	Date : 2007, 2000 Journal : *Libération*
2 **C**	88 ST 18.03 %	Fin (chi 2 : 31,14), mannequin, indice_de_masse_corporelle, perdre, long, mois, publicitaire, kilo, mort, proposer, magazine, pays, assurer, problème, commencer, poser, etc.	Date : 1999 Journal : *L'Humanité*
3 **B**	110 ST 22.54 %	Trouble (chi 2 : 86,08), alimentaire, comportement, mental, anorexie, maladie, boulimie, pathologie, anorexique, patient, pour_cent, site, psychique, boulimique, souffrir, toucher, etc.	Date : 2001, 2003, 2008 Journal : *La Croix*
4	91 ST 18.65 %	Lieu (chi 2 : 43,24), familial, psychiatrie, nourriture, repas, suivre, institut_mutualiste_montsouris, adulte, adolescent, professeur, pari, philippe_jeammet, table, quitter, exprimer, discussion, façon, partager, etc.	Date : 1997, 1998, 2003 Journal : *Le Figaro*
5 (la plus différenciée) **A**	69 ST 14.14 %	Publicité (chi 2 : 118,83), charte, marque, oliviero_toscani, média, vêtement, ministre_de_la_santé, campagne, professionnel, image, signer, engagement, nu, no_li_ta, décharner, diversité, roselyne_bachelot, mode, corps, ministère_de_la_santé, volontaire, photographe, italien, égide, réaliser, parrainer, benetton, jean_pierre_poulain, conduite, rendre, texte, podium, isabelle_caro, groupe, mesure, valorisation, démarche, sociologue, promouvoir, privilégier, positif, organisation, choquer, esthétique, corporel, public, groupe, etc.	Date : 2007 Journal : *Le Monde*

L'étude comparée de ces trois tableaux montre que certains mondes lexicaux sont stabilisés; autrement dit, ils se retrouvent quasiment à l'identique quel que soit le paramétrage choisi, tandis que d'autres sont plus fluctuants. Nous pouvons également remarquer que le fait de diminuer le nombre de classes de la phase terminale (tableau 2) a réduit le nombre de mondes lexicaux et fait disparaître la classe 5 liée à la « prévention-règlementation de

l'anorexie » (tableau 1). Concernant le tableau 3, les opérations d'homogénéisation sémantique semblent avoir bousculé les classifications précédentes : apparition de la classe 4 et modification des formes de certaines autres classes. De façon générale, nous pouvons déjà écrire que les choix techniques et intellectuels ont un impact sur les résultats obtenus. Nous pouvons supposer que la taille du corpus joue un rôle dans ces disparités : plus le corpus est « petit », plus les différences seraient visibles.

Si nous regardons plus précisément les formes représentatives associées à ces différents mondes lexicaux, nous pouvons nommer la classe A (n° 5 dans le tableau 1 et n° 5 dans le tableau 3) « prévention-règlementation de l'anorexie ». Elle renvoie à la fois à la campagne publicitaire d'Oliviero Toscani en 2007 (*publicité*, *marque*, *oliviero_toscani*, *vêtement*, *nu*, *no_li_ta*, *décharner*, etc.) et à la signature d'une charte d'engagement sur l'image corporelle en 2008 (*charte*, *médias*, *ministre_de_la_santé*, *professionnel*, *image*, etc.). Deux segments de texte caractéristiques de ces classes peuvent l'illustrer[79] :

> « Cette **campagne** a été **réalisée** par le photographe **oliviero_toscani** pour no_l_ita, **marque** de **vêtements** inconnue, dont les affiches **montrent** le corps **nu** et **décharné** d'**isabelle_caro**, vingt_sept ans, un **mètre** soixante_cinq et trente_deux kilos[80] »[81];

> « une **charte** de bonne **conduite** contre l'**anorexie** des mannequins. Plutôt que d'**interdire** de défilés les mannequins trop maigres, le **ministère_de_la_santé** devrait **privilégier** une **démarche** de responsabilisation des **professionnels** de la **mode**[82] ».

Notons que cet univers thématique est associé aux modalités 2007 et 2008 de la variable « date » dans le tableau 1 et uniquement à celle de 2007 dans le tableau 3. Il est également associé au quotidien *Le Monde.* Cela ne signifie pas que les autres journaux n'évoquent pas ces événements mais que ces formes co-occurrentes

---

79 Les termes en gras correspondent aux formes représentatives de la classe.

80 « Publicité choc, en Italie, contre l'anorexie », *Le Monde*, 27 septembre 2007.

81 Pour une meilleure lisibilité, nous rétablissons la ponctuation, les apostrophes et les majuscules supprimées par le logiciel.

82 Pierre Bienvault, « Une charte de bonne conduite contre l'anorexie des mannequins », *La Croix*, 6 juillet 2007.

sont un peu plus présentes dans les discours du *Monde.* Ce monde lexical correspond au pic médiatique repéré avec Modalisa, autrement dit au moment où l'anorexie « entre » sur la scène médiatique. Remarquons toutefois que la proposition de loi visant à pénaliser l'incitation à l'anorexie (2008) n'apparaît pas dans cette classe A et ne figure dans aucune classe des trois tableaux[83]. Ce résultat est surprenant car plusieurs brèves et articles sont consacrés à cet événement dans notre corpus. Une analyse des fréquences des formes montre que le terme « proposition » est employé à 19 reprises dans les discours contre 17 pour le mot « charte », 15 pour « oliviero_toscani » et « campagne » ou encore 8 pour « signer ». Certes, la fréquence d'un mot ne signifie pas qu'il est nécessairement co-occurrent d'autres formes de façon significative mais, d'après notre connaissance des discours, il existe bien un univers lexical caractéristique de cet événement. Nous faisons donc l'hypothèse que l'absence des mots relatifs à cette décision politique dans les classifications obtenues s'expliquerait par une surface rédactionnelle peu importante de l'événement par rapport à l'ensemble du corpus.

Les classes B correspondent au deuxième monde lexical qui renvoie à la « définition de l'anorexie ». Les formes qui lui sont associées sont des synonymes de la maladie (*trouble*, *comportement*, *pathologie*...) et des qualificatifs la désignant (*alimentaire*, *mental*, *psychique*). Elles apparaissent dans les trois tableaux. En revanche, certains termes ne figurent pas dans tous les profils de classe. Ainsi, xavier_darcos est une forme très significative de la classe 2 du tableau 1 (chi 2 : 21,85). Il apparaît aussi dans la classe 3 du tableau 2 mais avec un chi 2 relativement faible (5,3) et cette forme est absente du tableau 3 (classe 3). Autre exemple : dans les discours médicaux sur l'anorexie, la maladie est souvent associée à la boulimie, ces pathologies étant considérées comme les deux facettes d'un même trouble du comportement alimentaire.

83 Les termes « proposition » et « loi » ne font pas partie des formes représentatives associées aux différentes classes. De plus, une vérification du corpus coloré en fonction des classes (option proposée par Iramuteq) confirme que la majeure partie des articles évoquant cet événement figurent dans les segments de texte du corpus non classés.

Le terme « boulimie » est effectivement fortement associé au monde lexical « définition de l'anorexie » dans le tableau 3 (chi 2 : 30,13). En revanche, elle est classée dans les formes non significatives du tableau 2 (chi 2 : 2,23) et apparaît dans le tableau 1 comme étant une forme peu significative (chi 2 : 8,13). Un segment de texte illustre cette co-occurrence des formes « anorexie » et « boulimie » dans le tableau 3 : « Quatre_vingt_dix **pour_cent** des **anorexiques** sont des femmes, tandis_que la **boulimie** touche soixante_dix à quatre_vingt **pour_cent** de femmes et vingt à trente **pour_cent** d'hommes. Des **troubles** complexes. L'**anorexie** et la **boulimie** sont les **expressions** les plus caractérisées des **troubles** alimentaires[84] ». La classe « définition de l'anorexie » se caractérise donc par quelques formes stables autour desquelles « gravitent » d'autres formes qui ne sont pas les mêmes dans les trois classifications hiérarchiques descendantes. Notre corpus se compose donc d'un ensemble de discours (ou de fragments de discours) qui définissent l'anorexie mentale notamment en fournissant des indications épidémiologiques. Toutefois, les frontières de cette thématique fluctuent en fonction des paramétrages choisis et des modifications sémantiques apportées au corpus.

Les mêmes remarques peuvent être formulées quant à la classe C. Quelques formes stables la caractérisent : *kilo*, *mannequin*, *poids*, *magazine*, *perdre*. Le recours aux segments de texte spécifiques de ce monde lexical nous permet de mieux en saisir la thématique : « **Régimes** : la chair est triste. Dès le retour du printemps, haro sur les **kilos.** "Pour **perdre** du **poids**, un **régime** tu entreprendras", tel est le credo qui se décline à l'infini sur les pages glacées des **magazines**[85] ». Il s'agit ici d'une critique de la dictature de la minceur imposée par la presse magazine féminine, présente notamment dans les discours de *L'Humanité*. En effet, dans les trois classifications, cet univers thématique est associé à

84 « De grands couturiers s'inquiètent de la course à la maigreur », *La Croix*, 13 janvier 2007.

85 Michel Clerget, « Régimes : la chair est triste », *L'Humanité*, 1er avril 1999.

ce quotidien, de façon plus ou moins forte[86]. En revanche, notons que les modalités de la variable « date » ne sont pas toujours les mêmes. Comme cela a été expliqué précédemment, nous supposons qu'en modifiant le paramétrage et certains éléments du corpus, les frontières des segments de texte au sein desquels sont repérées les formes co-occurrentes changent, tout comme les corrélations entre formes co-occurrentes. Par conséquent, les fragments du corpus regroupés pour former les classes le sont de façon légèrement différente[87]. Notons que l'association de la modalité *L'Humanité* à ce monde lexical – et plus généralement la présence de cette thématique dans la classification obtenue – est surprenante au vu du faible nombre d'articles de ce quotidien dans notre corpus (3 sur 39).

Si l'on prête attention aux autres mots de la classe C, nous pouvons remarquer que le terme « top_modèle » apparaît dans les deux premiers tableaux mais est absent du troisième. La modalité 2006 de la variable « date » est également associée à ces deux classes. La présence de cette forme peut signifier que des fragments de discours portant sur le décès d'un mannequin anorexique à l'été 2006 sont intégrés à ce monde lexical dans les deux premières classifications. Un segment de texte caractéristique de cet univers thématique l'illustre : « La faim tragique d'une **top_modèle** au brésil. Ana_carolina_reston, quarante **kilos** pour un **mètre** soixante_quatorze, est **décédée** des **suites** de son anorexie. La **mort** d'une **top_modèle** brésilienne, mardi, des **suites** de son anorexie[88] ». L'adjectif « brésilien » – qui fait référence à la nationalité du mannequin – figure également dans la liste des formes représentatives de ces deux classifications. Nous

86 Les chi 2 associés à la variable sont plus ou moins significatifs : 7,51 pour le tableau 1, 13,89 pour le tableau 2 et 6,19 pour le tableau 3.

87 En effet, comme le soulignent Ludovic Lebart et André Salem, « le découpage du corpus en partie revêt une importance primordiale pour la construction d'une classification à partir de l'ensemble des formes d'un même corpus car les variations dans la partition du corpus ont pour effet de rapprocher certains couples de formes et d'en éloigner d'autres ce qui influe forcément sur les classes de la hiérarchie », Ludovic Lebart et André Salem, *op. cit.*, p. 127.

88 AFP et Reuters, « La faim tragique d'une top model au Brésil », *Libération*, 17 novembre 2006.

pouvons conclure sur la difficulté à nommer ce monde lexical puisqu'il renvoie d'un côté à la dictature de la minceur imposée par la presse magazine, dessinant implicitement un facteur déclencheur socioculturel de l'anorexie, et de l'autre, à un événement qui a suscité une polémique sur les exigences du métier de mannequin.

Enfin, la classe D comporte peu de formes considérées comme très significatives[89], leur chi 2 étant relativement faible en comparaison avec celui des formes de la classe A par exemple (chi 2 > 100). Cela signifie que ce monde lexical se démarque moins des autres univers thématiques du corpus : il est peu spécifique. Un segment de texte caractéristique, extrait d'un article de *Libération* paru en 2007, nous donne des indications sur la thématique dont il est question : « **Mardi** vingt_six **juin**, quatorze **heures** dix : le défilé des maux d'adolescents. Comme tous les **mardis** après_midi, **marcel_rufo**, le plus **médiatique** des pédopsychiatres, **consulte**. Il **aime** ce **moment**. L'**homme** est charmant, bavard[90] ». Ce fragment de texte est extrait d'un article portant sur une émission télévisée consacrée à la Maison de Solenn, alors dirigée par Marcel Rufo. Le lecteur est placé en situation de spectateur des consultations du psychiatre, filmées pour le reportage. Il semble que cette classe D s'organise principalement autour de cet article, relativement long, et qui est le seul du corpus à évoquer cette émission. Par conséquent, cela peut expliquer la faible spécificité de cette classe.

Au terme de ces analyses, nous pouvons écrire qu'il n'y a pas de « temporalité thématique » clairement identifiable dans nos trois classifications à l'exception de l'année 2007, voire 2008, relatives à la « prévention-règlementation de l'anorexie ». Le discours sur l'anorexie semble plutôt diffus, les modalités de la variable « date » étant rarement identiques et souvent non consécutives. En outre, nous avons vu avec l'analyse de la classe C qu'un même monde lexical peut renvoyer à des éléments distincts (facteur déclencheur socioculturel de l'anorexie / décès

89 Dont le seuil de signification du chi 2 est inférieur à 0,0001.

90 Éric Favereau, « Dans les mots de Rufo », *Libération*, 5 juillet 2007.

de mannequins anorexiques) qui demandent un retour au contexte pour être précisés. Concernant les divergences pointées entre les différentes classifications, nous supposons que la taille de notre corpus peut constituer l'une des causes des écarts soulignés. Toutefois, même avec un corpus volumineux, si certains mots des discours – quantitativement très fréquents – sont amenés à être modifiés[91], les mondes lexicaux évolueront eux aussi. L'approche en termes de mondes lexicaux ne permet donc pas forcément de valider ou d'infirmer des hypothèses précises mais plutôt de disposer d'un premier aperçu du contenu d'un corpus pour formuler ou affiner des hypothèses de recherche[92].

Conclusion : Quel(s) logiciel(s) pour quel(s) usage(s)?

Notre analyse ne prétend pas à l'exhaustivité mais entendait donner un aperçu de certaines fonctionnalités de deux logiciels d'analyse automatisée de discours. Au terme de cette contribution, nous souhaitons conclure plus largement sur la question du choix des logiciels dans un travail de recherche. Ce choix dépend des objectifs scientifiques, de la taille du corpus à traiter, des connaissances dont dispose le chercheur en statistique, mais également d'une sensibilité personnelle. En effet, il semble évident que des corpus de plusieurs centaines – voire milliers – d'articles ne peuvent être traités entièrement « à la main » au risque, pour le chercheur, de ne plus voir certains éléments et/ou de laisser ses intuitions et ses *a priori* l'influencer. Concernant la dimension statistique, les logiciels tels que Modalisa (ou du moins l'usage que nous en avons) sont relativement simples à manier et s'appuient sur des éléments de statistique descriptive (histogrammes, tri croisé et à plat), facilement appropriables. En

91 Comme nous l'avons fait pour le terme « anorectique ».

92 Christophe Lejeune rappelle que les concepteurs de ces outils, comme Max Reinert, « leur assignent avant tout un rôle heuristique (aider la formulation d'hypothèses) et non probatoire » (Christophe Lejeune, *op. cit.*, p. 24). En effet, pour Max Reinert, « ce ne sont pas des instruments de validation mais des aides à la construction d'hypothèses, ou même plus simplement, des aides à la lecture, car on doit recomposer soi-même sa tresse du sens » (Max Reinert, « La tresse du sens et la méthode "Alceste"... », *op. cit.*, p. 10).

revanche, les logiciels d'« analyse automatique » comme Iramuteq nécessitent des connaissances précises en statistique textuelle à la fois pour s'approprier le vocabulaire analytique et pour interpréter les résultats. Comme le soulignent Ludovic Lebart et André Salem, les « règles d'interprétation des représentations obtenues par le biais de ces techniques de réduction[93] n'ont pas la simplicité de celles de la statistique descriptive élémentaire[94] ». L'acquisition de ces connaissances n'est pas toujours évidente et peut décourager les chercheurs novices. Elle doit également s'accompagner d'une « vigilance épistémologique[95] » car les procédés statistiques sur lesquels reposent ces logiciels ne sont pas neutres. Le chercheur ne peut donc faire l'économie d'une réflexion sur les postulats théoriques qui les sous-tendent afin de comprendre ce que « produit » le logiciel. Cette compréhension n'est pas toujours aisée mais nous semble indispensable pour un usage raisonné et pertinent de ces outils. Enfin, au-delà des aspects scientifiques et techniques, c'est aussi une question de sensibilité personnelle qui peut pousser le chercheur à intégrer ces outils à son cadre méthodologique. En effet, les logiciels d'« analyse automatique » et « manuelle » n'appréhendent pas le discours de la même façon. Un logiciel d'« analyse automatique » « découpe des unités de la chaîne textuelle pour réaliser des comptages utilisables par les analyses statistiques ultérieures[96] », ces unités étant ensuite recontextualisables de différentes manières. Lors de cette opération, les mots sont réduits à des formes graphiques dépourvues de sens[97]. À l'inverse, le codage d'un corpus dans Modalisa – logiciel d'« analyse manuelle » – repose sur la compréhension sémantique que le chercheur a des discours ainsi que sur sa connaissance du sujet. Cela lui permet de prendre en compte des éléments et des mots qui sont peu nombreux dans le corpus, mais pertinents pour la recherche, éléments qui risquent d'être invisibilisés par une approche en termes de mondes lexicaux comme le propose

93 Les méthodes factorielles et les méthodes de classification « automatique ».
94 Ludovic Lebart et André Salem, *op. cit.*, p. 80.
95 Christophe Lejeune, *op. cit.*, p. 17.
96 Ludovic Lebart et André Salem, *op. cit.*, p. 8.
97 Par exemple, le logiciel ne comprend pas les métaphores et l'ironie.

Iramuteq. De ce fait, l'approche « automatique » d'un corpus peut ne pas convenir aux chercheurs attachés à la singularité des discours et des mots.

Bibliographie

AFP et Reuters, « La faim tragique d'une top model au Brésil », *Libération*, 17 novembre 2006.

Aubert, Aurélie, « Analyse de contenu et statistiques : une étude de la prise de parole des téléspectateurs au travers de leurs courriels », dans Camille Laville, Laurence Leveneur et Aude Rouger (dir.), *Construire son parcours de thèse – Manuel réflexif et pratique*, Paris, L'Harmattan, 2008, p. 97-104.

Azeddine, Leila, Gersende Blanchard et Cécile Poncin, « Le cancer dans la presse d'information générale. Quelle place pour les malades? », *Questions de communication*, n° 11, 2007, p. 111-127.

Bardin, Laurence, *L'analyse de contenu*, Paris, Presses Universitaires de France, 1985.

Bienvault, Pierre, « Une charte de bonne conduite contre l'anorexie des mannequins », *La Croix*, 6 juillet 2007.

Blandin, Claire *et al.*, « Femmes et alcool dans la presse écrite française (1980-2012) : construction sociale des problèmes de santé publique et des rapports de genre », *Les cahiers de l'IREB*, n° 21, 2013, p. 153-159, http://www.ireb.com/sites/default/files/Cahiers%2021.pdf, site consulté le 17 septembre 2015.

Charaudeau, Patrick (dir.), *La médiatisation de la science. Clonage, OGM, manipulations génétiques*, Bruxelles, Éditions de Boeck, 2008.

Clerget, Michel, « Régimes : la chair est triste », *L'Humanité*, 1er avril 1999.

Dalud-Vincent, Monique, « Alceste comme outil de traitement d'entretiens semi-directifs : essai et critiques pour un usage en sociologie », *Langage et société*, n° 135, 2011, p. 9-28.

Delavigne, Valérie, « Alceste, un logiciel d'analyse textuelle », *Texto! Textes et cultures, Équipe Sémantique des textes*, 2003, https://hal.archives-ouvertes.fr/hal-00924168/document, site consulté le 21 septembre 2015.

Delforce, Bernard, « La responsabilité sociale du journalisme : donner du sens », *Les Cahiers du journalisme*, n° 2, 1996, p. 16-32.

Delforce, Bernard, « Discursivité sociale / discours sociaux : penser les enjeux sociaux de l'information », dans Aurélie Tavernier *et al.* (dir.), *Figures sociales des discours. Le "discours social" en perspectives*, Lille, Éditions des Presses de l'Université Charles-de-Gaulle Lille 3, 2010, p. 57-72.

Delforce, Bernard et Jacques Noyer, « Pour une approche interdisciplinaire des phénomènes de médiatisation : constructivisme et discursivité sociale », *Études de communication*, n° 22, 1999, p. 14-37.

Ducos, Alexia *et al.*, « Classification d'un corpus hétérogène : la page Facebook de soutien au "bijoutier de Nice" (septembre 2013) », 12e Journées internationales d'Analyse statistique de Données Textuelles, 2014, p. 225-238, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/18-JADT2014.pdf, site consulté le 21 septembre 2015.

Favereau, Éric, « Dans les mots de Rufo », *Libération*, 5 juillet 2007.

Garcin-Marrou, Isabelle, *Des violences et des médias*, Paris, L'Harmattan, 2007.

Greimas, Algirdas-Julien, *Du sens II : essais sémiotiques*, Paris, Seuil, 1983.

Jacquez, Lise, « La controverse autour des expulsions des sans-papiers dans la presse française : une analyse des discours et des enjeux sociopolitiques », thèse de doctorat en sciences de l'information et de la communication, Lyon, Université Lyon 2, 2014, http://theses.univ-lyon2.fr/documents/lyon2/2014/jacquez_l/pdfAmont/jacquez_l_these.pdf, site consulté le 21 septembre 2015.

Joselin, Laurence *et al.*, « Dynamiques temporelles de la pandémie de grippe A/H1N1 dans la presse écrite francophone », 12e Journées internationales d'Analyse statistique de Données Textuelles, 2014, p. 299-310, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/25-JADT2014.pdf, consulté le 21 septembre 2015.

Lasègue, Charles, « De l'anorexie hystérique », *Journal français de psychiatrie*, n° 32, 2009, p. 3-8.

Laügt, Olivier, « Le SRAS dans *Le Monde* : un agent double? », *Les Cahiers du journalisme*, n° 15, 2006, p. 86-101.

Lebart, Ludovic et André Salem, *Statistique textuelle*, Paris, Dunod, 1994.

Lejeune, Christophe, « Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative », *Recherches qualitatives*, n° 9, 2010, p. 15-32.

Loubère, Lucie, « Le traitement des TICE dans les discours politiques et dans la presse », 12e Journées internationales d'Analyse des Données

Textuelles, 2014, p. 433-445, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/36-JADT2014.pdf, consulté le 21 septembre 2015.

Mayaffre, Damon, « Plaidoyer en faveur de l'Analyse de Données co(n) Textuelles. Parcours cooccurrentiels dans le discours présidentiel français (1958-2014) », 12e Journées Internationales d'Analyse statistique des Données Textuelles, 2014, http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/01-JADT2014.pdf, consulté le 21 septembre 2015.

Mayaffre, Damon, « Les corpus *réflexifs* : entre architextualité et hypertextualité », *Corpus*, n° 1, 2002, http://corpus.revues.org/11, consulté le 2 décembre 2014.

Mayaffre, Damon, « Analyses lexicologiques et rhétoriques du discours », 1er-11 décembre 2009, Alexandrie, http://eprints.aidenligne-francais-universite.auf.org/19/1/pdf_Formation_Mayaffre_Alexandrie_dec09_.pdf, consulté le 21 septembre 2015.

Moirand, Sophie, « Du traitement différent de l'intertexte selon les genres convoqués dans les événements scientifiques à caractère politique », *Semen*, n° 13, 2001, p. 97-117.

Neveu, Erik, « L'approche constructiviste des "problèmes publics" – Un aperçu des travaux anglo-saxons », *Études de communication*, n° 22, 1999, p. 41-57.

Noyer, Jacques, « La couverture du sida dans la presse française de 1982 à 1989 à travers trois quotidiens nationaux (*Le Figaro*, *Libération*, *Le Monde*) : approches de la notion d'événement », thèse de doctorat en sciences de l'information et de la communication, Lille, Université Lille 3, 1994.

Pincemin, Bénédicte *et al.*, « Concordanciers : thème et variations », 2006, https://halshs.archives-ouvertes.fr/halshs-00154100/file/pincemin_al_jadt06_texte.pdf, consulté le 21 septembre 2015.

Ratinaud, Pierre et Pascal Marchand, « Recherche improbable d'une homogène diversité : le débat sur l'identité nationale », *Langages*, 2012, n° 187, p. 93-107.

Reinert, Max, « Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars », *Langage et société*, n° 66, 1993, p. 5-39.

Reinert, Max, « La tresse du sens et la méthode "Alceste". Applications aux "Rêveries du promeneur solitaire" », 5e Journées internationales d'Analyse Statistique des Données Textuelles, 2000, http://lexicometrica.univ-paris3.fr/jadt/jadt2000/pdf/31/31.pdf, site consulté le 21 septembre 2015.

Reinert, Max, « La méthode d'analyse exploratoire des données textuelles "Alceste" et le problème de l'analyse de contenu », *Les Cahiers de Jérico-st*, n° 4, Presses de l'Université de Tours, 2004, p. 79-90.

Reinert, Max, « Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours », *Langage et société*, n° 121-122, 2007, p. 189-202.

Roche, Émilie, « Étude des discours de presse écrite française sur la violence et la torture pendant la guerre d'Algérie : *Le Monde*, *L'Humanité*, *Le Figaro*, *L'Express*, *France-Observateur*, 1954-1962 », thèse de doctorat en sciences de l'information et de la communication, Lyon, Université Lyon 2, 2007, http://theses.univ-lyon2.fr/documents/lyon2/2007/roche_e#p=0&a=top, site consulté le 21 septembre 2015.

Roy, Normand et Roseline Garon, « Étude comparative des logiciels d'aide à l'analyse de données qualitatives : de l'approche automatique à l'approche manuelle », *Recherches qualitatives*, vol. 32, n° 1, p. 154-180.

Touboul, Annelise *et al.*, « La disparité des modes de traitement journalistiques et des énonciations éditoriales sur le Web. Le cas d'un sondage sur Marine Le Pen et la Présidentielle de 2012 », *Réseaux*, n° 176, 2012/6, p. 73-103.

« De grands couturiers s'inquiètent de la course à la maigreur », *La Croix*, 13 janvier 2007.

« Publicité choc, en Italie, contre l'anorexie », *Le Monde*, 27 septembre 2007.

Sitographie

http://www.inserm.fr/ : site Internet de l'INSERM

http://www.iramuteq.org/ : site Internet d'Iramuteq

http://www.modalisa.com/index.php : site Internet de Modalisa