

Analyse des apports de la lexicométrie comme source de triangulation de représentations sociales : corroboration, complémentarité et contradiction

Daniel Pélissier¹

¹Université Toulouse 1 Capitole – daniel.pelissier@iut-rodez.fr

Abstract

This article presents a comparative analysis of a thematic analysis and a lexicometric approach for the study of social representations. The studied corpus gathers speeches of young graduates commenting on their own browsing on the Internet within the framework of a recruitment in the banking sector. The contributions of Reinert's hierarchical top-down classification are of three types: a reinforcement of the results of the thematic analysis, complementarities and contradictions. This last category highlights the ability of lexicometry to promote the researcher's reflexivity.

Keywords: comparison, contribution, triangulation, reflexivity, social representations.

Résumé

Cet article présente une analyse comparative d'une analyse thématique et d'une approche lexicométrique pour l'étude de représentations sociales. Le corpus étudié regroupe des discours de jeunes diplômés commentant leur propre navigation sur internet dans le cadre d'un recrutement dans le secteur bancaire. Les apports de la classification hiérarchique descendante de Reinert sont de trois ordres : une corroboration des résultats de l'analyse thématique, des complémentarités et des contradictions. Cette dernière catégorie souligne l'intérêt de la lexicométrie comme source de réflexivité du chercheur.

Mots clés : comparaison, apport, triangulation, réflexivité, représentations sociales.

1. Introduction

Repérer et analyser une représentation sociale est une opération complexe tant cet objet est difficile à définir (Moscovici, 1988). Plusieurs courants de recherche s'intéressent à cette problématique (Moliner et Guimelli, 2015), chacun adoptant des méthodes différentes.

Dans l'étude présentée dans cette proposition, nous aborderons l'intérêt de croiser les regards sur ces représentations sociales en utilisant des analyses thématique et lexicométrique. La lexicométrie n'échappe pas au postulat que toute méthode a ses propres biais. L'argumentation de son usage a pu être « *enthousiaste* » (Pincemin, 2012, p. 18) ou critique (Carbou, 2017 a). Plutôt que proposer une analyse de son apport *in abstracto*, nous préférons dans cet article faire émerger son intérêt par la comparaison de résultats de recherche issus de deux méthodes, une analyse thématique et une classification descendante hiérarchique (Reinert, 1983). Cette approche comparative tend à souligner des complémentarités entre méthodes (Desmarais et Moscarola, 2004, Martin et al, 2016) et non à déterminer la supériorité d'une approche sur l'autre.

Après avoir synthétisé le contexte de la recherche, nous présenterons les résultats qui sont de trois types : des corroborations des analyses qualitatives, des compléments et contradictions.

2. Présentation synthétique de la recherche et problématique méthodologique

Notre analyse de l'apport de la lexicométrie à l'analyse de représentations sociales sera réalisée à partir d'une recherche précise. Ce cas sera présenté dans cette section de façon synthétique pour faire apparaître les choix méthodologiques principaux.

2.1. Contenu et contexte du recueil des discours analysés

La recherche concernait l'étude de la réception de traces numériques de banques par de jeunes diplômés dans un contexte de recrutement. Ces discours ont été collectés par la méthode de l'autoconfrontation simple (Pélissier, 2017) qui consiste à faire commenter sa navigation par un utilisateur *a posteriori*. 21 jeunes diplômés ont été rencontrés en 2016. Ils ont dans un premier temps navigué sur internet sur des sites relatifs à une banque française. Ils ont ensuite commenté cette navigation auprès du chercheur. Les discours analysés dans la suite de cet article correspondent à ces commentaires de navigation (fig. 1). Cette méthode ne repose pas sur un guide d'entretien mais laisse libre le jeune diplômé à la fois dans sa navigation et son analyse du parcours sur internet. Les discours recueillis sont composés d'environ 70000 occurrences pour 180 pages de retranscription. Plus précisément, ce corpus représente 67261 occurrences, 2797 formes, 78,12 occurrences par intervention de locuteurs en moyenne.

Retranscription	Corpus Iramuteq avec variables
<p>[>Question?]: pouvez-vous m'expliquer votre navigation ?</p> <p>[>3IL1]: ce que j'ai remarqué déjà, ce petit encart sur les indices boursiers. Ça a l'air de rien mais c'est quand même significatif de la santé d'une entreprise. Première chose que je fais, toujours histoire du groupe [scroll bas rapide].</p>	<p>**** *Num_3IL1 *ProfilFB_oui *FrequenceFB_unefoisparjour *RSpro_oui *FrequenceRSPro_plusieursfoisparmois *RechercheWeb_ouitoujours *InfluenceWeb_ouiselon *Abandon_non *CandidatureWeb_oui *RSent_oui *SiteTemoign_oui *CandidatBanque_eventuellement *TravailBanque_oui *RechercheEmploi_nonouvert *Age_23 *Sexe_M *Nationalite_francaise *MetierPere_ONQ *MetierMere_employeAdmin *Groupe_3IL</p> <p>Ce que j'ai remarqué déjà, ce petit encart sur les indices boursiers. _Ca_ a l'air de rien mais c'est quand même significatif de la santé d'une entreprise. Première chose que je fais, toujours histoire du groupe.</p>

Figure 1. Extrait du corpus analysé

2.2 La lexicométrie comme source de triangulation

L'hypothèse centrale de cette recherche argumentait que les discours des jeunes diplômés étaient orientés par des représentations sociales. Comme cet objet est complexe et difficile à repérer, il a été décidé de croiser plusieurs méthodes d'analyse.

Une première analyse a consisté à repérer qualitativement des représentations sociales par une analyse thématique. Pour améliorer la fiabilité des résultats, un processus de triangulation des méthodes (Denzin, 1978) a été mis en œuvre. La triangulation peut se définir comme « *l'emploi d'une combinaison de méthodes et de perspectives permettant de tirer des conclusions valables à partir d'un même phénomène* » (Fortin, 1996, p. 318). La lexicométrie a été retenue pour trianguler les résultats de l'analyse qualitative par une construction de classes, assez indépendante des choix du chercheur (Desmarais et Moscarola, 2004).

Notre problématique méthodologique est ainsi de comprendre quels sont les apports de la lexicométrie comme outil de triangulation d'une analyse qualitative. Avant de présenter les

principaux résultats, nous détaillerons la mise en œuvre des deux méthodes triangulées, l'analyse thématique et la classification descendante hiérarchique.

2.3 Représentations sociales et analyse croisée des discours

Deux catégories de méthodes ont été retenues pour repérer des représentations sociales dans les discours de ces jeunes diplômés.

2.3.1. L'analyse thématique

L'analyse thématique a consisté, dans cette recherche, au « *repérage, au regroupement et, subsidiairement, à l'examen discursif des thèmes abordés dans un corpus* » (Paillé et Mucchielli, 2012, p. 232). En ce sens, elle se distingue de l'analyse catégorielle (Campenhoutt et Quivy, 2011, p. 208) qui vise à quantifier par leur durée ou le nombre de mots des ensembles de discours analysés comme sémantiquement proches. Nous avons adopté une approche qualitative centrée sur le contenu et les relations entre les thèmes.

Deux démarches sont possibles pour l'établissement des thèmes : une approche continue qui part de l'analyse du discours pour construire tous les thèmes ou une analyse séquentielle qui préfère établir les thèmes *a priori* (Paillé et Mucchielli, 2012, p. 237). Nous avons commencé par établir une liste de thèmes en fonction de nos hypothèses de recherche ce qui favorisait une analyse cohérente avec notre problématique. Puis nous avons adapté, modifié, étendu cette liste au fil de la retranscription avec le logiciel SONAL (Alber, 2010) ce qui favorise les découvertes du terrain. La liste ainsi actualisée après chaque entretien a permis aussi d'avoir un indicateur de saturation des thèmes.

Une attention particulière a été accordée aux contenus des thèmes qui, dans la mesure du possible, renvoyaient au discours des acteurs et à leur vécu (Lejeune, 2016, p. 65). Cette démarche réflexive a été suivie dans un journal de thématization¹ (Paillé et Mucchielli, 2012, p. 269). Le résultat de cette analyse thématique correspond à des arbres thématiques ou des cas. Chaque cas est organisé selon la structure définie par Guimelli (1995) qui distingue dans les représentations sociales des connaissances (description de l'objet), des prescriptions (conseils d'usages par rapport à l'objet) et des jugements (évaluation de l'objet).

Par exemple, les sites web comprennent les branches suivantes codifiées dans SONAL :

Branche CONNAISSANCES (6) : connaissances_eval_structure ; connaissances_eval_contenu ; connaissances_eval_moteurrecherche ; connaissances_image_entreprise ; connaissances_image_technique ; connaissances_image_partisane. Branche USAGES (2) : usages_postuler ; usages_entretien. Branche JUGEMENTS (5) : jugements_positifs_propre ; jugements_positifs_professionnel ; jugements_positifs_detaille ; jugements_negatifs_charge ; jugements_negatifs_recherche.

Les précautions prises pour mettre en œuvre cette analyse thématique ne suppriment pas les limites de cette démarche, notamment le choix des thèmes par le chercheur selon ses propres

¹ Ce journal correspondait à un tableau par entretien avec les colonnes suivantes : Numéro, Prénom ; saturation : quantitatif, qualitatif : les nouveaux thèmes ou l'approfondissement de thèmes existants ; synthèse des informations et titre : un titre et une synthèse des principales informations étaient rédigés ; suivi des hypothèses : RS (représentations sociales) taille, RS Facebook et témoignages : d'éventuelles validations même partielles de nos hypothèses étaient précisées ; profils : des profils étaient établis à chaud ; commentaires et journal de bord : des commentaires ouverts ont été précisés dans cette colonne ainsi que les éléments principaux du journal de bord manuscrit ; passages intéressants : les contenus qui nous paraissaient les plus pertinents étaient relevés ; navigation : les caractéristiques de la navigation sont précisées dans cette colonne.

représentations. Aussi, nous avons triangulé l'analyse thématique avec une approche lexicométrique.

2.3.2. La classification descendante hiérarchique

La classification utilisée notamment par la lexicométrie est une « *technique statistique permettant de regrouper des individus ou observations entre lesquels a été définie une distance* » (Lebart et Salem, 1988, p. 180). M. Reinert (1983) a développé une approche particulière de Classification Descendante Hiérarchique (C.D.H.) qui part du corpus global vers des classes de plus en plus précises. La démarche débute par un découpage en segments de texte qui peuvent correspondre à tout ou partie de phrases, de paragraphes. Puis le chercheur peut décider de retenir uniquement les formes pleines et lemmatiser les formes. Une fois ces étapes réalisées, la C.D.H. (Ratinaud et Marchand, 2012, p. 837) est mise en œuvre.

Les classes ainsi obtenues peuvent être associées à des représentations sociales. Les mots utilisés dans un discours écrit ou oral renvoient notamment à des appartenances sociales (Marchand, 2007, p. 53), à des groupes particuliers, des environnements, etc. Le langage est ainsi imprégné de social et nos choix de vocabulaire ne sont pas neutres ou anecdotiques. Cette relation a conduit M. Reinert à proposer la notion de '*mondes lexicaux*' qui sont « *la trace d'usages ou 'lieux communs', ceux-ci étant prégnants à la fois dans des discours particuliers (ceux soumis à l'analyse) et dans des manières sociales de penser* » (Reinert, 2007, p. 195). Les classes obtenues par la C.D.H. rassemblent des formes présentes dans des segments de texte et l'homogénéité de ces rassemblements renverrait à des mondes lexicaux, par nature, proches des représentations sociales. Ainsi, « *il y aurait une organisation normative des mots, qui permettrait d'investir des lexiques pour montrer notre appartenance groupale* » (Ratinaud et Marchand, 2015, p. 59). De plus, l'usage de variables externes (âge, sexe, lieu, etc.) permet d'associer une classe à un contexte et ainsi aborder les dynamiques représentationnelles (Geka et Dargentas, 2010, p. 115). Cette relation supposée entre classes de la C.D.H. et représentations sociales a été testée empiriquement (Moliner et al. 2002 ; Kalampalikis, 2003 ; Geka et Dargentas, 2010 ; Ratinaud et Marchand, 2015 ; Bouchard et Kasparian, 2018) et a montré qu'il pouvait exister un lien entre l'usage de formes et une représentation du monde, représentation partagée en raison de la fréquence d'usage de ces formes.

Nous avons effectué plusieurs C.D.H. simples sur segments de texte avec le logiciel *Iramuteq* (Ratinaud, 2009) en faisant varier le nombre de classes demandées et le nombre minimum de segments par classe. Nous avons choisi de retenir les formes dont la fréquence est supérieure à 3 (soit 1056 formes dans ce corpus) pour centrer le traitement sur les formes les plus fréquentes. Le retour au texte permet de nuancer cette décision en remarquant certaines formes plus rares. Au terme de ces simulations, nous avons retenu une C.D.H. qui présente 16 classes avec un taux de segments classés de 86,16 % (fig. 2) qui est un taux très satisfaisant (Peyrat-Guillard, 2014, p. 522).

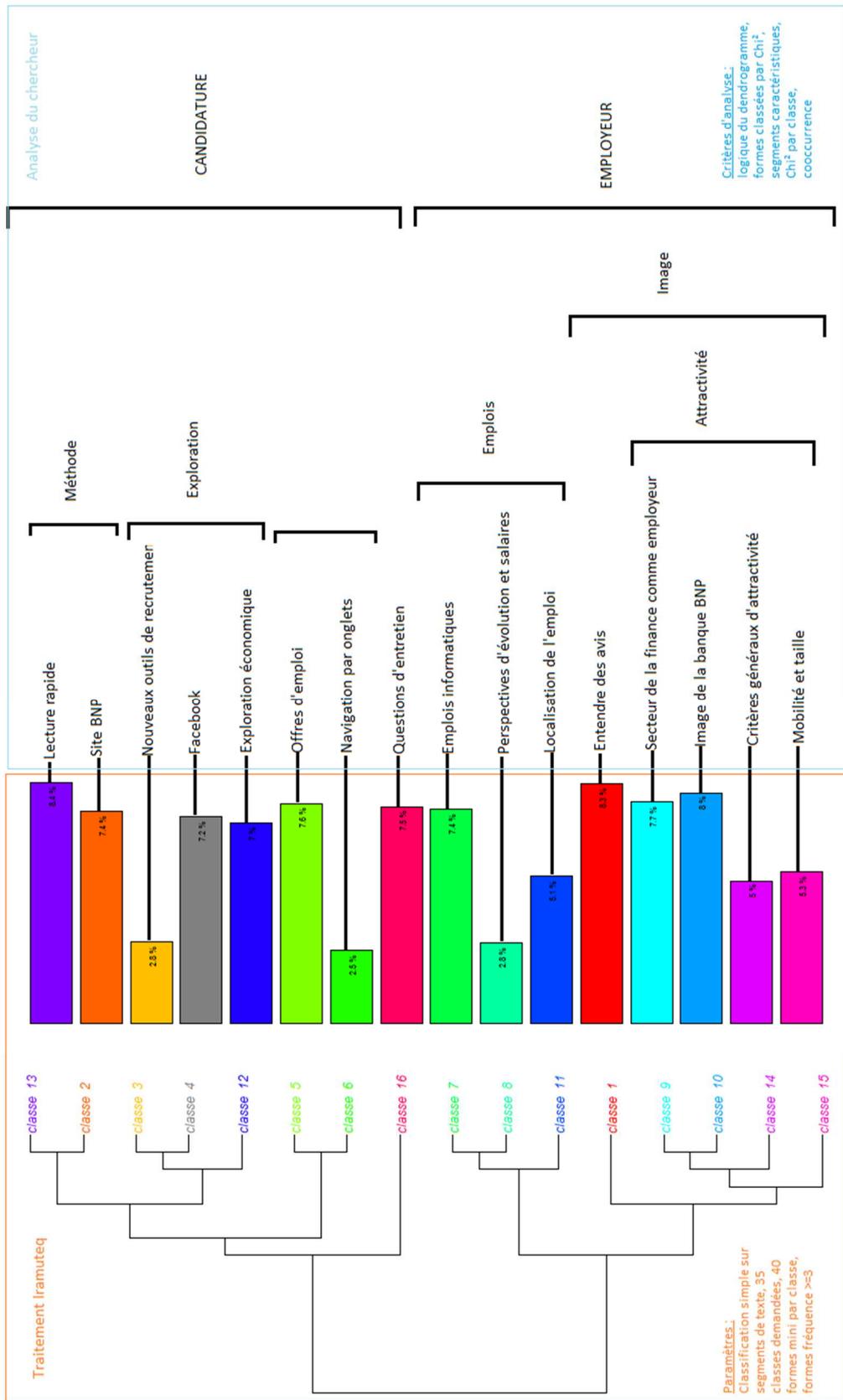


Figure 2. Classification descendante hiérarchique des entretiens d'autoconfrontation et analyse des résultats

3. Résultats : les trois apports de la lexicométrie comme source de triangulation

L'analyse thématique a permis de repérer plusieurs représentations sociales dans les discours de ces jeunes diplômés. Un premier ensemble concernait l'entreprise de façon générale et plus précisément sa taille et son appartenance au secteur bancaire. Le deuxième groupe de représentations sociales analysées à partir de l'analyse thématique se rattachait aux dispositifs de communication : les sites internet, la page Facebook et les plateformes d'avis de salariés. Selon l'analyse thématique, les commentaires de ces jeunes diplômés sur leur navigation étaient orientés par des représentations sociales à la fois de l'organisation au cœur de leur discours, la banque, et aussi par les dispositifs de communication, supports des informations observées.

Ce résultat a été confronté à une analyse lexicométrique pour croiser le regard sur ce corpus. Trois types d'apport de cette triangulation seront détaillés : des corroborations des résultats de l'analyse qualitative, des compléments et enfin des contradictions soulevées par certains résultats.

3.1 La corroboration du repérage de certaines représentations sociales

Certains résultats sont communs à l'analyse thématique qualitative et à l'analyse lexicométrique.

Les entretiens d'autoconfrontation laissent apparaître une classe associée à la taille de l'organisation. La classe 15 (fig. 2) présente en effet des segments de texte qui renvoient à l'évocation de cette représentation sociale pour expliquer la réception de ces jeunes diplômés. L'extrait suivant est ainsi caractéristique² de cette classe : « *mais par contre, peut-être les inconvénients d'une grande structure, le système hiérarchique pas démodé mais trop cloisonné du fait que ce soit une grande entreprise aussi du coup peut-être un peu moins de considération, d'après ce que j'ai lu sur les avis, voilà* ». Le raisonnement se base sur la grandeur de l'entreprise pour déduire des connaissances sur cette banque. Dans ce corpus, la forme « *grand* » est alors souvent associée à « *structure* »³ et à « *groupe* »⁴.

De même, la représentation sociale de Facebook est repérée par la méthode de la C.D.H. La classe 4 (fig. 2) regroupe 7,2 % des segments classés. La méthode lexicométrique utilisée met aussi en évidence la classe 1 regroupant des discours partagés sur les avis de salariés (fig. 2). Les segments repérés renvoient clairement à cette thématique. Enfin, la classe 9 (7,7% des segments classés) correspond à la représentation sociale de la banque dans un contexte d'emploi.

La taille de l'organisation, les pages Facebook, les plateformes d'avis de salariés et la banque correspondent ainsi à des représentations sociales avec l'approche lexicométrique. La triangulation permet la « *corroboration* » de résultats (Miles et Huberman, 2013, p. 481) plus qu'une validation. Ce résultat corrobore ainsi les conclusions de l'analyse thématique en soulignant une interprétation identique avec deux méthodes différentes ce qui accroît la fiabilité des interprétations du chercheur.

3.2 Les compléments au repérage de certaines représentations sociales

Certaines analyses issues de l'approche lexicométrique ont complété celles réalisées avec l'analyse thématique telle qu'elle a été mise en œuvre dans cette recherche. Deux types

2 Somme des Chi²=232,88.

3 Chi²=21,43 et cooccurrence=4.

4 Chi²=5,16 et cooccurrence=5.

d'apports ont été repérés : une amélioration de la compréhension du contenu du discours et une mise en évidence de relations statistiques avec certaines variables.

3.2.1. La compréhension du discours

La mise en œuvre de la méthode de la C.D.H. a favorisé une compréhension complémentaire du discours de ces jeunes diplômés. Nous citerons quelques exemples significatifs.

La représentation sociale de la taille est précisée par cette analyse lexicométrique. Elle est associée au thème de la mobilité⁵ dans ce corpus. Les formes associées à cette classe renvoient ainsi aux problématiques de mobilité « *interne* »⁶ de « *possibilité* »⁷ d'évolution de mobilité « *internationale* »⁸ ou encore de « *carrière* »⁹. Cette récurrence souligne le raisonnement selon lequel une « *grande* » banque a des ressources et une taille favorisant cette mobilité.

De plus, le jugement est plutôt positif¹⁰ pour ces grandes structures, avec l'usage d'adjectifs comme « *dynamique* »¹¹ ou « *intéressant* »¹². La forme « *dynamique* » est associée à « *jeune* »¹³.

De même, la représentation des plateformes d'avis est précisée par l'usage du vocabulaire utilisé qui navigue entre « *commentaire* » et « *avis* » dans les discours recueillis. La forme est très majoritairement « *avis* »¹⁴ mais il existe quelques usages de « *commentaires* »¹⁵.

Ces trois exemples illustrent des situations où la lexicométrie apporte un complément d'interprétation à l'analyse qualitative à partir des fréquences de formes associées aux classes renvoyant à des représentations sociales.

Un autre apport a découlé de la compréhension de la construction des classes. La lecture du dendrogramme (fig. 2, gauche ou bas) explique en effet la méthode suivie par l'algorithme et cette analyse permet de replacer le discours d'une classe dans un ensemble plus large. Le cas de la représentation sociale de la taille est représentatif de l'apport de cette démarche.

Le dendrogramme (fig. 2) commence par les classes concernant l'image et les avis (classes 15, 14, 10, 9, 1) puis celles sur l'emploi (11, 8, 7). Dans les classes relatives à l'image employeur, la classe 1 (avis) se détache en raison de son homogénéité, selon les critères de cette méthode. D'après ces discours, un employeur fournit des emplois et une partie des classes concerne ce thème. Les autres classes évoquent principalement la mobilité (classe 15), l'image en général (classe 10), le secteur et la banque (classe 9). La taille intervient dans plusieurs classes : la mobilité principalement (classe 15) et aussi le secteur financier qui est associé aux « *grandes* » banques (classe 9). Ainsi, les discours sur la taille relèvent plus d'une problématique d'analyse de l'employeur que de la préparation de la candidature.

5 Chi²=144,29, fréquence=15.

6 Chi²=123,69.

7 Chi²=60,02.

8 Chi²=51,92.

9 Chi²=24,77.

¹⁰ Cette analyse est nuancée en raison des limites et biais interprétatifs de l'orientation dite 'positive' d'une forme isolée de son contexte. Dans ce cas, l'interprétation a été réalisée à partir de la lecture des segments de textes et des textes dont étaient issues les formes associées à des discours dits 'positifs'.

11 Chi²=39,93.

12 Chi²=11,42.

13 Chi²=31,87 et cooccurrence=4.

14 Fréquence=59 dans cette classe.

15 Fréquence=18 dans cette classe.

Contrairement à l'analyse thématique, cette organisation du corpus est réalisée sans les *a priori* du chercheur (Blanchet et Gotman, 2007, p. 102). En effet, ce dernier intervient peu¹⁶ dans la détermination des classes qui sont formées par les différentes itérations réalisées. L'intérêt est d'avoir un regard différent sur les discours avec une logique autre que celle déduite de la problématique et hypothèses de recherche. Pour autant, cette approche ne remplace pas le regard du chercheur et n'offre pas une analyse de contenu toute faite (Reinert, 2007 b, p. 39). En effet, la C.D.H. prépare simplement l'analyse qui dépend toujours de la subjectivité du chercheur (Kalampalikis, 2003, p. 163 ; Marchand, 2007, p. 49 ; Geka et Dargentas, 2010, p. 129). Comprendre la construction des classes par l'algorithme est ainsi une opportunité de découvertes de compléments à l'analyse thématique.

3.2.2. Les relations statistiques entre discours et variables

Un autre intérêt de l'approche lexicométrique dans cette recherche a été de repérer des relations statistiques entre les discours et certaines variables. Chaque locuteur était associé à plusieurs variables issues des résultats d'un questionnaire posé avant la navigation internet (fig. 1 pour la liste des variables utilisées). L'association entre une variable et une classe mesurée par un test du Chi² a permis de montrer que certains discours comme ceux sur la taille de l'entreprise (classe 15) étaient plus liés à certains candidats (fig. 3).

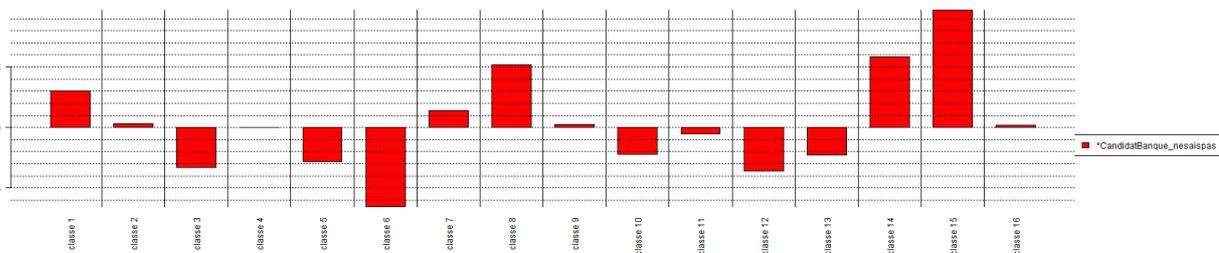


Figure 3. Chi² par classe et dendrogramme, variable Candidature Banque = « ne sais pas »

Les candidats ne sachant pas s'ils souhaitent candidater dans une banque sont fortement liés à la classe 15 mettant en évidence une représentation sociale de la taille (fig. 3). Par contre, ceux étant éventuellement prêts à candidater dans une banque ne sont pas associés à cette classe (fig. 4).

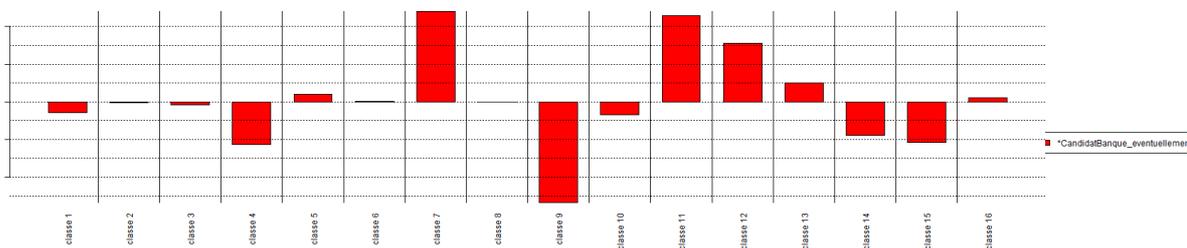


Figure 4. Chi² par classe et dendrogramme, variable Candidature Banque = « éventuellement »

Cette relation statistique relativise les interprétations sur la représentation sociale de la taille d'une organisation qui apparaît avec l'analyse thématique et aussi la C.D.H.

¹⁶ Le chercheur peut intervenir sur le choix du nombre de classes minimum, le nombre et le type des formes retenues, le choix de la lemmatisation, etc. Une fois ces choix effectués, la classification est réalisée par l'algorithme de l'application utilisée.

Ces deux premiers apports, corroboration et compléments, ont été mis en évidence dans la littérature (Paindorge, 2015, Martin et al. 2016) qui démontre l'intérêt des complémentarités entre méthodes (Desmarais et Moscarola, 2004, Martin et al, 2016). Dans ce cas, les résultats de la lexicométrie seraient en accord avec l'analyse qualitative ou complémentaires. Dans notre recherche, certains résultats sont contradictoires entre les deux approches. La prochaine section détaillera certaines de ces contradictions.

3.3 Les résultats contradictoires de la lexicométrie

La lexicométrie offre la possibilité de contredire les résultats obtenus avec d'autres approches. Toujours dans un postulat de non hiérarchisation des approches, les résultats contradictoires ne seront pas des remises en cause d'une des deux approches mais plutôt une source de réflexion sur les méthodes suivies. La contradiction est ainsi un des apports de la triangulation (Miles et Huberman, 2013, p. 483) aux tactiques de vérification des résultats.

Pour cette recherche, une contradiction est apparue : la représentation sociale des sites internet repérée avec l'analyse thématique n'apparaissent pas dans la C.D.H.

Le champ lexical du site internet est dilué dans des classes plus larges. La classe 2 regroupe des segments de texte sur le « *site* » internet de « BNP Paribas ». Elle est l'une des dernières classes construites de la catégorie des classes sur la candidature (fig. 2). Elle s'intègre à une logique de l'exploration par lecture rapide appliquée à un cas particulier, celui du site internet de BNP Paribas. Ce site internet est associé aux témoignages de salariés, et plus précisément au site *Glassdoor* intégré dans la page d'accueil de l'espace carrière. La problématique générale est cependant plus l'exploration et ses méthodes que la fiabilité des informations. La classe 4 sur Facebook présente aussi des segments de texte contenant la forme « *site* ». Les *verbatim* de cette classe associent alors le site dit « *officiel* » avec la page Facebook par les liens hypertextes utilisés dans les *posts* ou certains commentaires. Comme pour la classe 2, la logique est celle de l'exploration d'un dispositif numérique qui ne présente pas un champ sémantique spécifique dans ce corpus.

La méthode de l'autoconfrontation qui consiste à décrire une navigation a favorisé l'utilisation de formes liées à l'exploration des discours plus qu'à des débats sur les dispositifs de communication. Cependant, l'existence de classes homogènes de formes pour les autres dispositifs étudiés, Facebook et les « *avis* » de salariés, montre une situation particulière du site internet, au-delà d'un contexte méthodologique particulier.

Cette analyse contredit ainsi la nature de la représentation sociale du site internet pourtant repérée dans l'analyse qualitative des entretiens d'autoconfrontation. L'interprétation du site internet ne répond pas à la même logique que la page Facebook ou les avis de salariés. Si cette classification lexicométrique ne suffit pas pour conclure à l'absence d'une représentation sociale du site internet, elle dénote des discours et des processus cognitifs différents selon les dispositifs pour ces jeunes diplômés. Orienté vers la préparation de la candidature, le site internet a renvoyé vers des problématiques plus larges, comme la fiabilité ou les méthodes d'exploration d'une page internet, alors que la page Facebook et les « *avis* » ont provoqué des analyses plus homogènes. Le site internet semble échapper au débat, à l'interrogation de l'internaute comme s'il faisait partie d'un décor devenu invisible par la force de l'habitude. Les pages Facebook consacrées au recrutement ou les plateformes d'avis de salariés ont étonné ces jeunes diplômés, peu habitué à ces dispositifs, ce qui a pu provoquer des discours ciblés et homogènes ressortant dans la classification. Le site internet peut ainsi être une représentation masquée qui, sans être contre normative, se retrouve dans une zone muette (Abric, 2003) à la faveur d'une méthode particulière.

Cette situation contradictoire questionne les conditions de la recherche. En effet, le lien entre C.D.H. et représentation sociale a été établi par plusieurs études mentionnées précédemment. Le repérage de représentation sociale par une analyse thématique est aussi une approche connue. Les deux méthodes devraient déboucher sur le même constat ce qui est le cas pour 4 représentations sur les 5 repérées. Pourquoi cette différence d'analyses ?

L'analyse thématique a été questionnée dans un premier temps. La construction de thèmes a pu être influencée par un système d'hypothèses construit au préalable. Cette approche a pu conduire à survaloriser certains discours pour faire correspondre les hypothèses au corpus. La récurrence des discours est difficile à établir avec une approche qualitative et les associations statistiques repérées par l'approche lexicométrique peuvent être masquées dans l'analyse thématique. L'approche retenue n'a pas quantifié les thèmes car « *le nombre de répétitions ne définit pas l'importance d'un thème* » (Paillé et Mucchielli, 2012, p. 266). Or, la fréquence est un indicateur du partage de la représentation et de son caractère social. Mais, cette contradiction questionne aussi la construction des classes. Si la C.D.H. repère plus des mondes lexicaux que des thèmes (Carbou, 2017 b, p. 9-11), une relation trop directe entre classe et représentation sociale peut être délicate pour un objet aussi complexe. Les résultats communs obtenus pour 4 représentations ne permettent pas de conclure de façon catégorique à une représentation sociale mais argumentent une hypothèse forte qui devra être explorée par d'autres méthodes sur d'autres terrains. De plus, la classification analysée n'est pas seule possible et d'autres pourraient isoler des discours plus centrés sur les sites. Au-delà des méthodes retenues pour trianguler nos analyses, les conditions de la recherche ont aussi été questionnées : ordre des analyses, nombre de chercheurs, connaissances des méthodes, etc.

Finalement, les contradictions mises en évidence par la triangulation nous guident vers une « *série d'explication plus complexe* » (Miles et Huberman, 2013, p. 482). La contradiction soulignée par la triangulation de la lexicométrie questionne le chercheur ce qui est une façon de développer sa réflexivité qui « *apprécie le point de vue du chercheur et de sa relation avec le terrain comme source décisive de données et d'interprétation* » (Kuehner et al., 2016, p. 700).

Dans cette recherche, la contradiction portant sur la représentation sociale du site internet conduit à nuancer la conclusion de l'analyse qualitative. Elle interroge aussi sur la méthode classificatoire utilisée pour repérer les autres représentations sociales. Elle replace l'analyse obtenue dans un contexte particulier de recueil ayant influencé la compréhension du contenu des discours.

4. Conclusion

Cet article a souhaité souligner les apports de la lexicométrie comme source de triangulation de l'analyse de discours de jeunes diplômés. Ces derniers sont de nature différente : corroboration, complémentarité mais aussi contradiction. Les résultats présentés ne prétendent pas à l'exhaustivité pour cette recherche. Ils montrent plutôt pour chacune de ces catégories des contributions possibles de la lexicométrie à un processus de recherche souhaitant croiser les regards. La possibilité de résultats contradictoires met en évidence l'apport réflexif de l'approche lexicométrique, notamment pour les chercheurs isolés et les petits corpus.

Cette approche triangulatoire a aussi ses propres limites. L'usage de l'analyse thématique est multiple tout comme celle de la C.D.H. Même en réalisant une comparaison sur des discours identiques, les éléments soulignés ne peuvent prétendre à une généralisation. Il est probable que d'autres chercheurs, avec ces mêmes discours, auraient repéré des éléments sensiblement différents. La lexicométrie, dans le cas étudié, a permis d'analyser la subjectivité du chercheur, souligner certaines limites de chaque méthode et remettre en contexte des résultats sur un objet de recherche protéiforme, les représentations sociales. Corroborant et complétant l'analyse

thématique qualitative, elle a aussi apporté des contradictions propres à améliorer la compréhension des biais et des limites des résultats obtenus.

La comparaison proposée dans cet article est finalement celle de deux constructions. L'analyse thématique présentée s'attache à la compréhension du discours des locuteurs par leurs représentations partagées grâce à la construction de thèmes ou de catégories. Cette analyse est construite par un chercheur ayant ses propres représentations, sa subjectivité. Le croisement avec la lexicométrie n'a pas permis d'objectiver ces réflexions mais plutôt d'apporter un autre regard dont la construction diffère.

La cohérence épistémologique de cette comparaison repose en effet sur une vision constructiviste de la constitution des connaissances pour laquelle la compréhension du sujet connaissant est centrale. Un des postulats de cet article est que la lexicométrie peut s'intégrer à une épistémologie constructiviste dès lors qu'elle assume la subjectivité du chercheur. Que ce soit lors du choix de la méthode retenue, de sa mise en œuvre ou lors de l'analyse des résultats, le chercheur investissant la lexicométrie construit la connaissance d'un objet, des représentations sociales dans ce cas, qui n'est pas « donnée » par des traitements. La statistique textuelle n'évince pas la subjectivité du chercheur (Carbou, 2017a, p. 104) mais offre, dans cet article, l'opportunité de réfléchir sa pensée au miroir d'une présentation construite. La déconstruction de la démarche lexicométrique exige ainsi une analyse de la construction du corpus, des algorithmes utilisés et des interprétations du chercheur. Cet article a proposé, par une présentation contrastée de la lexicométrie, de questionner la possibilité d'un positionnement constructiviste de cette analyse textuelle fondée sur des statistiques. Le débat est complexe et mérite de nombreuses discussions et articles mieux centrés sur cette problématique de fond. L'aborder par un cas précis permet de limiter la tentation de généralisations simplificatrices.

Bibliographie

- Abric J.-C. (2003). La recherche du noyau central et de la zone muette des représentations sociales. Dans *Méthodes d'étude des représentations sociales* (Abric, J.-C., p. 59-80). Paris : ERES.
- Alber A. (2010). Voir le son : réflexions sur le traitement des entretiens enregistrés dans le logiciel Sonal. *Socio-logos*, 5.
- Bouchard M.-A. et Kasparian S. (2018). *La classification hiérarchique descendante pour l'analyse des représentations sociales dans une pétition antibilinguisme au Nouveau-Brunswick, Canada*. Communication présentée au Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles, Rome (p. 142-149).
- Campehouth L. V. et Quivy R. (2011). L'exploration. Dans *Manuel de recherche en sciences sociales* (Campehouth, L., Quivy, R., p. 107-138). Paris : Dunod.
- Carbou G. (2017a). Analyser les textes à l'ère des humanités numériques, quelques questions pour l'analyse statistique des données textuelles. *Les Cahiers du Numérique*, (3-4), 91-114.
- Carbou G. (2017b). Quelques questions à l'attention des utilisateurs de statistique textuelle pour l'analyse des discours. *Texte!*, XXII(4), 1-20.
- Denzin N. K. (1978). *The research act : a theoretical introduction to sociological methods*. New York : McGraw-Hill.
- Desmarais C. et Moscarola J. (2004). Analyse de contenu et analyse lexicale, Le cas d'une étude en management public. *Actes JADT*, 1-20.
- Fortin M.-F. (1996). *Le processus de recherche : de la conception à la réalisation*. Montréal : Décarie.
- Geka M. et Dargentas M. (2010). L'apport du logiciel Alceste à l'analyse des représentations sociales : l'exemple de deux études diachroniques. *Les Cahiers Internationaux de Psychologie Sociale*, I(85), 111-135.

- Guimelli C. (1995). Valence et structure des représentations sociales. *Bulletin de psychologie*, 49(422), 58-72.
- Kalampalikis N. (2003). L'apport de la méthode Alceste dans l'analyse des représentations sociales. Dans *Méthodes d'étude des représentations sociales* (Abric, J.-C., p. 147-163). Paris : Erès.
- Kuehner A., Ploder A. et Langer, P. C. (2016). Introduction to the Special Issue: European Contributions to Strong Reflexivity. *Qualitative Inquiry*, 22(9), 699–704.
- Lamalle C. et Salem A. (2002). *Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels*. Saint Malo (p. 403-411). Communication présentée au 6ème journées internationales d'analyse statistique des données textuelles.
- Lebart L. et Salem A. (1988). *Analyse statistique des données textuelles*. Paris : Dunod.
- Lejeune C. (2017). Analyser les contenus, les discours et les vécus, à chaque méthode ses logiciels ! Dans *Les méthodes qualitatives en psychologie et sciences humaines de la santé* (Santiago-Delefosse, M., del Rio Carral, M., p. 205-224). Paris : Dunod.
- Marchand P. (2007). Concepts, méthodes, outils. Dans *Analyse statistique de données textuelles en sciences de gestion* (Gauzente, C., Peyrat-Guillard, D., p. 49-70). Caen : EMS Management et société.
- Martin A. Adele, S. et Reutenauer C. (2016). Stratégies du voyageur : analyse croisée d'entretiens semi-directifs. *Actes JADT*, 1-12.
- Miles M. B. et Huberman A. M. (2013). *Analyse des données qualitatives* (Miles, M. B., Huberman, A. M). Bruxelles : de boeck.
- Moliner P. et Guimelli C. (2015). *Les représentations sociales*. Grenoble : PUG.
- Moliner P. Rateau P. et Cohen-Scali V. (2002). *Les représentations sociales, pratique des études de terrain* (Presse universitaire de Rennes).
- Moscovici S. (1988). Esquisse d'une description des représentations sociales. 1988. Dans *Le scandale de la pensée sociale* (Moscovici, S., p. 19-64). Paris : EHESS.
- Paillé P. et Mucchielli A. (2012). L'analyse qualitative en sciences humaines et sociales. Dans *L'analyse qualitative en sciences humaines et sociales* (p. 183-205). Paris : Armand Colin.
- Paindorge M., Kerneis J. et Fontanieu, V. (2015). Analyse de données textuelles informatisée : l'articulation de trois méthodologies, avantages et limites. *Revue internationale de systémique complexe et d'études relationnelles*, 11(1), 65-92.
- Pélissier D. (2017). Observer les observateurs, adaptation de la méthode de l'autoconfrontation simple et complexité de la navigation internet. *Recherches qualitatives, Hors-série*(22), 24-41.
- Peyrat-Guillard D., Lancelot Miltgen C. et Welcomer S. (2014). Analysing conversational data with computer-aided content analysis: The importance of data partitioning. *Journées internationales d'Analyse statistique des Données Textuelles*, 519-530.
- Pincemin B. (2012). Sémantique interprétative et textométrie. *Texto!*, XVII(3).
- Ratinaud P. (2009). Iramuteq [Linux, Windows, MacOS] (version 0.7 Alpha 2). Lerass.
- Ratinaud P. et Marchand P. (2012). *Application de la méthode ALCESTE aux « gros » corpus et stabilité des 'mondes lexicaux' : analyse du « CableGate » avec IRAMUTEQ*. Liège (p. 835-844). Communication présentée au Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles.
- Ratinaud P. et Marchand P. (2015). Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014). *Mots. Les langages du politique*, 108(2), 57-77.
- Reinert M. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, 8(2), 187-198.
- Reinert M. (2007). Contenu des discours et approche statistique. Dans *Analyse statistique des données textuelles en sciences de gestion* (Gauzente, C., Peyrat-Guillard, D., p. 23-45). Caen : EMS Management et société.