



HAL
open science

Comprendre les réactions aux réformes en Belgique francophone à travers les médias écrits grâce à l'analyse lexicométrique avec le logiciel IRaMuTeQ

Mélanie Ferrara, Nathanaël Friant

► To cite this version:

Mélanie Ferrara, Nathanaël Friant. Comprendre les réactions aux réformes en Belgique francophone à travers les médias écrits grâce à l'analyse lexicométrique avec le logiciel IRaMuTeQ . Journées internationales d'Analyse statistique des Données Textuelles 2016, UMR Bases, Corpus, Langage (CNRS – Université de Nice Sophia Antipolis)., Jun 2016, Nice, France. halshs-01329184

HAL Id: halshs-01329184

<https://shs.hal.science/halshs-01329184>

Submitted on 8 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comprendre les réactions aux réformes en Belgique francophone à travers les médias écrits grâce à l'analyse lexicométrique avec le logiciel IRaMuTeQ

Mélanie Ferrara¹, Nathanaël Friant²

¹Université de Mons – melanie.ferrara@umons.ac.be

²Université de Mons – nathanael.friant@umons.ac.be

Abstract

An attempt to estimate the gap between the normative and cognitive frameworks will be done through the analysis of press archives and parents-teacher associations' periodicals issued during periods during which major reforms occurred in the French-speaking Belgian educational system. In a first step, we highlighted a debates progression from a purely qualitative viewpoint (Ferrara & Friant, 2015). Although, the lexicometric approach seemed necessary to continue our exploratory research by the processing of a wider collection of documents that will enhance our first analysis. This work introduces the first outcomes of our textual data processing with IRaMuteQ through a top-down hierarchical clustering and a factorial correspondence analysis. Finally, we will compare this lexicometric approach with our former qualitative analysis in a critical perspective.

Résumé

Par l'analyse de corpus de presse et de périodiques d'associations de parents et d'enseignants sélectionnés autour de deux périodes qui ont connu des réformes majeures de l'enseignement secondaire en Belgique francophone, nous tenterons de mesurer les écarts entre les cadres cognitif et normatif inhérents à chaque période. Lors d'une première approche purement qualitative (Ferrara & Friant, 2015) nous avons pu mettre en évidence les premiers éléments d'une évolution des débats. Cependant, le recours à la lexicométrie nous a paru essentiel dans la poursuite de notre démarche exploratoire afin de pouvoir traiter un plus grand corpus et compléter notre première analyse. Ce travail reprend donc les premiers résultats obtenus par la classification hiérarchique descendante (CHD) et l'analyse factorielle des correspondances (AFC) sous IRaMuTeQ. Finalement, nous jetterons un regard critique sur cette approche lexicométrique en comparaison avec notre précédente analyse thématique.

Mots-clés : classification hiérarchique descendante, Analyse factorielle des correspondances, Iramuteq, réformes éducatives, presse, médias écrits.

1. Introduction

Nous avons travaillé sur la base d'articles de deux quotidiens nationaux belges francophones (Le Soir et La Libre Belgique) parus lors de deux périodes clés ayant connu la mise en place de réformes éducatives : la Rénovation de l'enseignement (1971) et le Contrat pour l'Ecole (2005) ainsi que d'articles de périodiques de deux associations de parents et d'une association d'enseignants. Lors d'une publication précédente (Ferrara & Friant, 2015), nous avons réalisé une analyse thématique ascendante en continu (Paillé & Mucchielli, 2012) d'articles parus en 1971 et en 2005, nous permettant d'identifier le passage d'un débat idéologique entre différentes conceptions de l'éducation en 1971 à un débat d'ordre technique en 2005. Cependant, l'analyse thématique exige un investissement en temps directement proportionnel à la taille du corpus, et deviendrait donc

impossible à mettre en œuvre sur des corpus beaucoup plus volumineux. Par conséquent, nous proposons dans cette communication une analyse exploratoire textuelle de ce corpus outillée au moyen du logiciel IRaMuTeQ (Ratinaud, 2009). L'analyse des médias écrits visera à rendre compte des représentations, des valeurs, des croyances qui ont été diffusées et mettra en évidence la manière dont le changement, la réforme ont été vécus et interprétés. Pour analyser ces corpus, nous ferons appel à des techniques de lexicométrie, à savoir la classification hiérarchique descendante (CHD) selon la méthode de Reinert (1983) et l'analyse factorielle des correspondances (AFC). L'AFC permettra d'organiser les données, après transformation statistique, sous forme de graphiques à deux dimensions faisant apparaître des relations lexicales entre elles. Ainsi, cette analyse nous permettra de mettre en lumière la distance entre les classes issues du corpus et donc quelles variables semblent s'opposer ou au contraire se rapprocher. La CHD, quant à elle, permettra de regrouper les « mondes symboliques lexicaux » communs. Ces différentes analyses favoriseront la mise en évidence précise et rapidement analysable des relations lexicales entre les différentes données (Marty, [Marchand & Ratinaud, 2013](#); Valence & Roussiau, 2013 ; Loubère, 2014 ; Reinert, 2008). Enfin, nous évaluerons, dans une perspective critique, la plus-value apportée par l'approche lexicométrique par rapport à notre précédente démarche qualitative.

2. Problématique

Aujourd'hui la dévalorisation des filières qualifiantes dans les représentations sociales des élèves en Belgique francophone est un fait avéré (Ferrara & Friant, 2014) et, comme nous le verrons, cette dévalorisation en tant que telle est au cœur des débats politiques en matière d'éducation depuis leur genèse. Si l'action réformatrice semble, à plusieurs reprises, avoir été considérée comme un passage obligé en vue de revaloriser ces filières et « lisser » les inégalités ; leur efficacité et les bénéfices qui en ont été tirés peuvent être remis en question à la lumière de la situation actuelle. L'une des pistes de recherche envisagée consiste en l'identification et l'évaluation de la dissonance entre, d'une part, la composante politique et normative du système éducatif de laquelle émanent les propositions de réforme et, d'autre part, ce que Lessard et Carpentier (2015) définissent comme le « cadre culturel cognitif » du système éducatif. C'est-à-dire la vision du monde des acteurs, leurs conceptions, ce qui les oriente vers des conduites qu'ils estiment légitimes. Ce concept de cadre culturel cognitif peut, à notre sens, être rapproché de celui de représentations sociales (Abric, 2003). Afin d'appréhender ce cadre culturel cognitif, nous avons choisi d'utiliser les médias écrits comme source de données. Afin de relever les courants et interactions sociales propres à chaque période, de mettre en évidence les jeux des acteurs et évaluer le « cadre cognitif » de chaque époque, nous avons choisi d'analyser des corpus de médias écrits pour les raisons évoquées précédemment mais aussi pour l'accessibilité plus ou moins aisée que constitue ce média. . Lors de travaux précédents (Ferrara & Friant, 2015), nous avons déjà soumis un corpus de presse à une analyse qualitative de type thématique ascendante à des fins de recherche similaires à celles qui nous motivent ici. Cependant, quelques manquements de résultats observés nous ont poussés à poursuivre notre travail exploratoire sur un corpus plus étoffé et en ayant recours à une analyse lexicométrique. Si le recours à une analyse de type lexicométrique à l'aide du logiciel Iramuteq nous a paru évidente pour des raisons d'économie de temps, nous espérons également mettre en exergue l'organisation et le contenu des discours des différents acteurs ainsi que la position inter-acteurs que nous n'avions pu déceler que de manière limitée par notre analyse thématique.

3. Les corpus

Nous avons travaillé sur la base de deux corpus distincts tous deux constitués d'archives de presse de deux journaux belges ainsi que d'archives de périodiques d'associations de parents et d'enseignants. Les corpus se distinguent par le type de décret étudié et sont donc tirés de deux périodes distinctes. Le premier est constitué d'extraits de textes traitant de la réforme du renové. Pour ce faire, nous avons sélectionné les articles de 1971 – année de mise en application de la réforme – jusqu'à 1973. Le second est constitué d'extraits de textes débattant du Contrat pour l'Ecole dont la date est comprise entre 2005 et 2007 pour les mêmes raisons que celles citées ci-dessus. Un nettoyage manuel des corpus s'est révélé nécessaire (Roy & Garon, 2013), principalement pour assurer l'homogénéité du texte. En effet, les décrets éducatifs contiennent des termes spécifiques qui perdraient leur sens s'ils étaient scindés et donc considérés comme deux formes distinctes. Certains mots ont donc été regroupés ou synthétisés (ex : renforcement des deux premières années du primaire = renforcementP1P2) afin d'en assurer la reconnaissance comme une seule unité sémantique de ceux-ci. Certains termes non pertinents pour l'analyse ont également été supprimés (ex : les noms des jours de la semaine, le prénom des Ministres, ...) afin de ne pas polluer les résultats. Finalement, nous obtenons pour le renové corpus composé de 183 textes dont 49583 occurrences, une moyenne des occurrences par texte de 270.95, 3729 formes et 1548 hapax (3.12%des occurrences - 41.51% des formes). Le corpus relatif au Contrat pour l'Ecole est lui, constitué de 197 textes dont 55207 occurrences, une moyenne d'occurrences par texte de 280.24, 4191 formes et 1771 hapax (3.21%des occurrences - 42.26% des formes).

4. Analyse

L'analyse que nous proposons ici, s'effectue en deux temps. Le premier temps se concentre sur les classifications des deux corpus afin de mettre en évidence les mondes lexicaux communs. Nous verrons s'il peut ressortir de leur interprétation qualitative une approche similaire à la thématisation. Ensuite, nous tenterons de mettre en évidence les relations entre acteurs en projetant les différents corpus sur deux dimensions grâce à l'AFC.

4.1. La classification hiérarchique descendante des corpus

Le tout premier temps de nos analyses consistait à effectuer une CHD sur chaque corpus d'après la méthode de Reinert (1983). Dans notre étude, nous avons choisi d'opérer cette analyse en classification simple sur segments de textes. Pour rappel, la CHD, comme son nom l'indique, opère une classification hiérarchique de type descendante qui, au départ du corpus unique de tous les segments, va distinguer itérativement des classes de segments homogènes selon une métrique prédéfinie, en l'occurrence la dissimilarité ou mesure de la distance inter-classes. En outre, la CHD ou toute autre classification hiérarchique peut-être perçue comme une mesure de l'entropie du corpus qui reflèterait dans notre cas la dynamique des débats pour chaque période étudiée (Jardino, 2014). Cette étape nous permettra d'obtenir des classes de mots dont la structure interne sera suffisamment cohérente pour pouvoir en extraire un sens interprétable.

4.1.1. CHD sur le corpus « Renové »

La CHD effectuée sur le corpus traitant de la réforme du renové a été initiée avec une demande de 15 classes terminales, le logiciel en a retenu 10 permettant ainsi de couvrir 86% de segments classés.

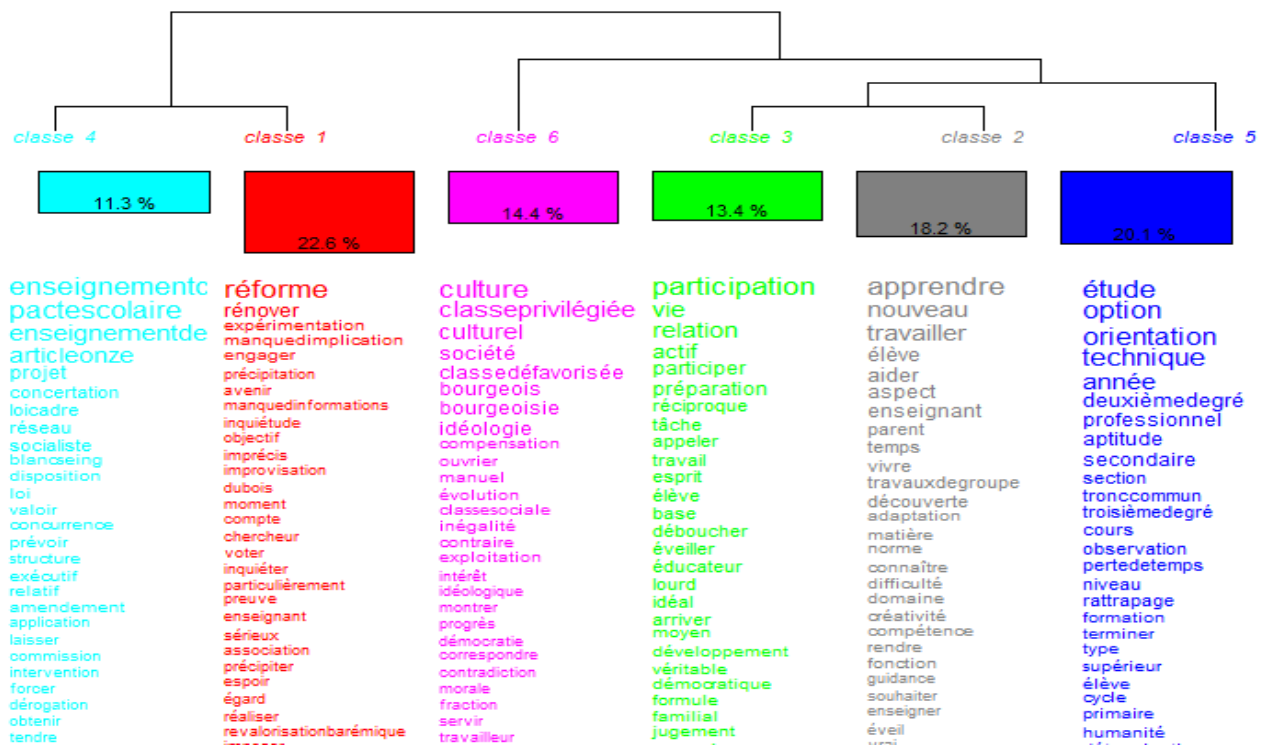


Figure 1. Dendrogramme de la CHD effectuée sur le corpus du rénové

Si l'on observe le dendrogramme dans son ensemble, on distingue deux branches principales avec, d'une part, les classes 4 et 1 et d'autre part, les classes 6, 3, 2, 5. La première branche étant davantage centrée sur des questions et débats de natures normatives ou organisationnelles tandis que la deuxième semble plutôt se rattacher aux implications organisationnelles de l'action réformatrice.

4.1.2. CHD sur le corpus « Contrat pour l'Ecole »

En ce qui concerne la CHD effectuée sur le corpus du Contrat pour l'Ecole, nous avons demandé 15 classes terminales, le logiciel en a retenu 10 permettant ainsi de couvrir 83,86% de segments classés.

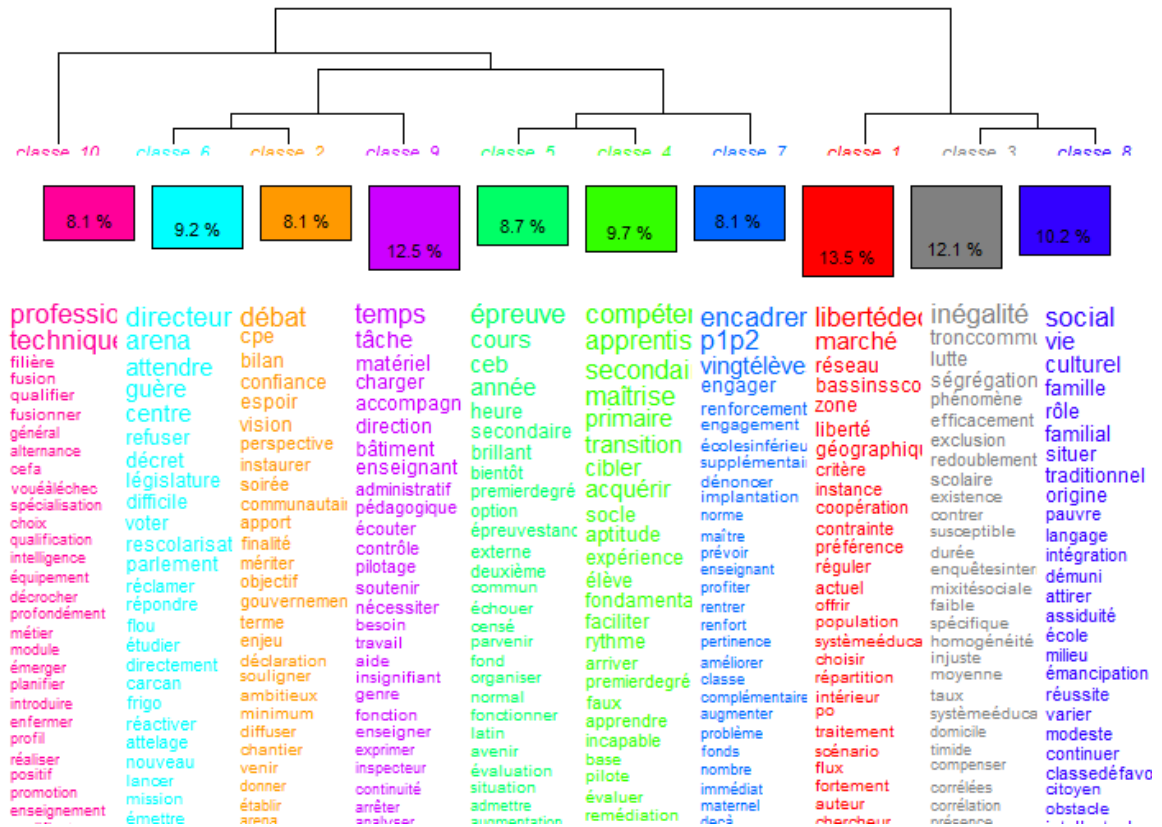


Figure 2. Dendrogramme de la CHD effectuée sur le corpus du Contrat pour l'Ecole

Encore une fois, si l'on remonte le dendrogramme, on voit que les dernières classes 1,3 et 8 sont assez rapidement mises en évidence. En effet, elles dépeignent clairement le caractère social et idéologique des discours représentant 35.8% du corpus sur seulement 3 classes. Les autres classes, quant à elles, sont moins homogènes et font plutôt référence aux aspects normatifs et organisationnels du corpus. Si l'on compare les résultats obtenus dans les deux CHD présentées ci-dessus, la diversité des classes et l'étendue du dendrogramme pour la période la plus récente du Contrat pour l'Ecole indiquent une plus grande dynamique des débats liés à cette réforme. En effet, cette période semble plus riche du point de vue des thèmes abordés. Cette dimension constitue un élément nouveau qui n'avait pas pu être dégagé lors de notre analyse thématique mais qui rejoint bien notre conclusion quant au niveau d'information des acteurs (qui animent les débats) d'une période par rapport à l'autre (Ferrara & Friant, 2015).

4.2. AFC sur la variable « acteurs »¹

¹ Les couleurs du graphique représentent les modalités de la variable acteur et sont identiques pour les figures 3 et 4 : Bleu : décideurs, rouge : chercheurs, gris : enseignants, vert : parents, turquoise : membres des partis de l'opposition et rose : syndicats.

Nous avons ensuite mené une AFC sur chacun des deux corpus en sélectionnant les formes actives de la variable « acteurs » suivant l'indice Chi2. La Figure 3 correspondant au corpus du rénové présente des sous-corpus plus ou moins distincts. Cependant, une classe de formes est dispersée de manière homogène sur tout le graphique, il s'agit des formes associées aux chercheurs en éducation. Cela peut d'une part s'expliquer par un niveau d'information et un discours plus étendu, mais aussi une absence de prise de position. Le facteur 2, qui fait ressortir l'empreinte normative du discours (du moins normatif en ordonnée positive au plus normatif en ordonnée négative) semble séparer les parents et les enseignants du monde politique constitué des décideurs, des syndicats et des partis de l'opposition. Le facteur 1 qui fait lui ressortir davantage la composante idéologique des débats (d'absence d'idéologie en abscisse négative à une forte empreinte idéologique en abscisse positive) marque une différence entre le discours des parents et celui des enseignants, ceux-ci se rejoignent cependant sur le plan normatif.

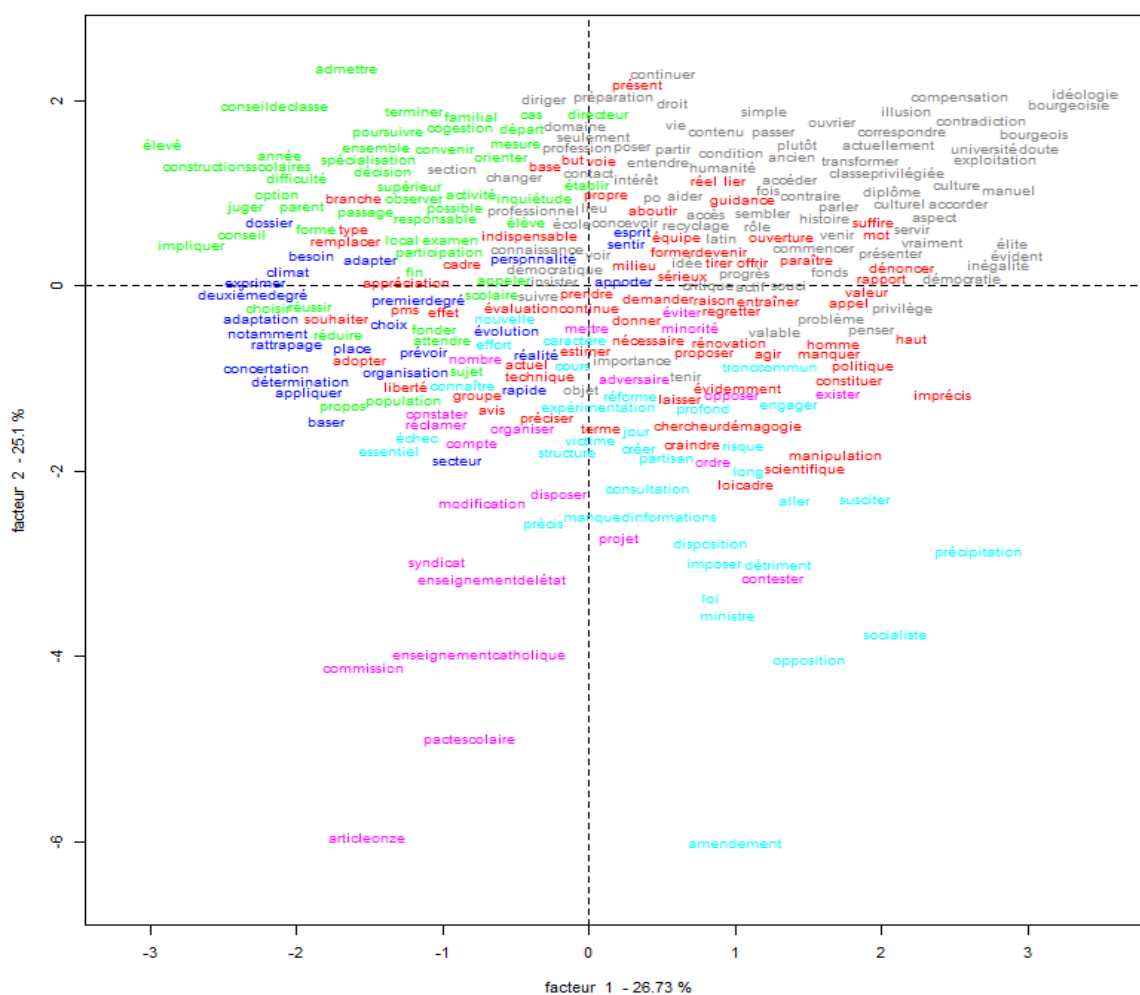


Figure 3. AFC sur le corpus "renové" suivant la variable "acteurs".

Sur la Figure 4 présentant les résultats de l'AFC sur le corpus du Contrat pour l'école on remarque une distribution différente. Les acteurs « politiques » (syndicats, décideurs, partis de l'opposition) sont opposés aux chercheurs, parents et enseignants suivant le facteur 1 qui correspond à une mesure de la teneur matérialiste du discours (d'exclusivement matériel en abscisse négative à majoritairement philosophique en abscisse positive). Enseignants et parents sont, cette fois, réunis, semblent partager des discours du même ordre et s'opposent

termes de possibilité de traitement et de projections visuelles, apporte une composante complémentaire intéressante à la démarche qualitative que nous avons initiée. De plus, la dimension heuristique des sorties logicielles constitue un outil réflexif supplémentaire pour le chercheur permettant d'accéder à des zones inexplorées du corpus par la démarche thématique pure. Par contre, l'économie de temps attendue initialement s'est révélée limitée. En effet, bien que ce type de logiciel permette de traiter des corpus importants, leur constitution – dans notre cas d'archives de médias écrits – représente la majeure partie du processus de recherche et même si le gain de temps pour la partie analytique est indéniable, il se retrouve quelque peu « noyé » dans le processus global. Finalement, il faut reconnaître un avantage à utiliser ce type d'analyse. Les logiciels associés, tel qu'IRaMuTeQ implémentent des méthodes éprouvées qui confèrent au chercheur un grand degré de liberté et d'objectivité dans son analyse et lui permettent de traiter des corpus de grande taille. Ceci sera nécessaire dans notre cas afin de mettre en évidence une évolutivité des discours lors de prochains travaux.

Références

- Abric, J. C. (Éd.). (2003). *Méthodes d'étude des représentations sociales*. Ramonville Saint-Agne: Erès.
- Ferrara, M., & Friant, N. (2014). Les représentations sociales des élèves du premier et du dernier degré de l'enseignement secondaire en Belgique francophone par rapport aux différentes filières. *L'orientation scolaire et professionnelle*, (43/4). <http://doi.org/10.4000/osp.4496>
- Ferrara, M., & Friant, N. (2015). Analyse socio-historique de réformes éducatives en Belgique francophone à travers la presse : Rénovation de l'enseignement (1971) et Contrat pour l'Ecole (2005). *Céreq, Relief* (50), 165-176.
- Jardino, M. (2004). Recherche de structures latentes dans des partitions de «textes» de 2 à K classes. *Présenté à 7èmes Journées internationales d'Analyse statistique des Données Textuelles*, Université Catholique de Louvain, Louvain-la-Neuve.
- Lessard, C., & Carpentier, A. (2015). *Politiques éducatives : la mise en œuvre*. Paris: Presses Universitaires de France.
- Loubère, L. (2014). Le traitement des TICE dans les discours politiques et dans la presse. *Présenté à 12èmes Journées internationales d'Analyse statistique des Données Textuelles*, Université de la Sorbonne nouvelle, Paris 3.
- Marty, E., Marchand, P., & Ratinaud, P. (2013). Les medias et l'opinion. *Eléments théoriques et méthodologiques pour une analyse du débat sur l'identité nationale*. *Bulletin de Méthodologie Sociologique*, 117(1), 46-60. <http://doi.org/10.1177/0759106312465550>.
- Paillé, P., & Mucchielli, A. (2012). *L'analyse qualitative en sciences humaines et sociales*. Paris: Armand Colin.
- Ratinaud P. (2009). *IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* [programme informatique]. <http://www.iramuteq.org>
- Reinert, M. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données, Vol VIII, 2*, 187-198
- Reinert, M. (2008). Mondes lexicaux stabilisés et analyse statistique de discours. *Présenté à 9èmes Journées internationales d'Analyse statistique des Données Textuelles*, Université de Versailles, Saint-Quentin-en-Yvelines

Roy, N., & Garon, R. (2013). Étude comparative des logiciels d'aide à l'analyse de données qualitatives: de l'approche automatique à l'approche manuelle. *Recherches qualitatives*, 32(1), 154-180.

Valence, A., & Roussiau, N. (2013). L'apport d'ALCESTE dans l'analyse médiatique des représentations sociales des droits de l'Homme et de l'immigration. *Présenté à Journées d'étude « Usages de la lexicométrie en sociologie »*, Université de Versailles, Saint-Quentin-en-Yvelines.

Van Haecht, A. (1985). *L'enseignement rénové: de l'origine à l'éclipse*. Bruxelles: Institut de sociologie. Editions de l'Université de Bruxelles.