

Hors thème

Étude comparative des logiciels d'aide à l'analyse de données qualitatives : de l'approche automatique à l'approche manuelle

Normand Roy, Ph.D.

Université du Québec à Trois-Rivières

Roseline Garon, Ph.D.

Université de Montréal

Résumé

L'objectif de cet article vise à effectuer un survol descriptif des plus récentes versions de douze logiciels d'aide à l'analyse de données qualitatives (LAADQ). La recension est organisée autour d'une typologie divisée en trois axes, soit les logiciels qui privilégient l'une ou l'autre des approches automatique, semi-automatique ou manuelle. Ces axes ne s'excluent pas mutuellement puisque de plus en plus de concepteurs combinent plusieurs modules dans un même logiciel, permettant de varier les analyses. Néanmoins, chaque logiciel propose habituellement une finalité prévalente, sur laquelle nous insistons lors de l'analyse comparative. Nous décrivons d'abord les caractéristiques principales des logiciels, puis nous présentons leurs forces et leurs limites. Un tableau synthèse comparatif à la fin de l'article permet une consultation rapide des différents outils disponibles sur le marché. Nous espérons ainsi pouvoir orienter les différents choix s'offrant aux chercheurs et faciliter leur décision.

Mots clés

MÉTHODE D'ANALYSE, LOGICIELS, ANALYSE QUALITATIVE, ORDINATEUR

Introduction

La recherche qualitative constitue un paradigme présent depuis un bon moment en sciences sociales. Il faut revenir aux années vingt pour retracer les bases de la recherche qualitative, établies, pour plusieurs sociologues, par le courant de l'École de Chicago (Coulon, 1992). C'est toutefois à partir des années soixante-dix que la recherche qualitative obtient ses lettres de noblesse aux

États-Unis et quelques années plus tard au Canada. Enfin, il faut attendre le début des années quatre-vingt-dix pour voir émerger les premiers LAADQ¹ (QSR International, 2012; ResearchWare, 2012; Scientific Software, 2012). Depuis ce temps, on peut constater un engouement grandissant envers l'utilisation de ces logiciels, en témoignent la publication de versions de plus en plus performantes et la multiplication de ces logiciels, dont des logiciels libres de droits. Par exemple, ATLAS.ti vient de faire paraître la version 7, alors que NVivo en est à sa dixième version. L'évolution et la multiplication constantes des outils rendent très difficile la tâche de se maintenir à jour, ce qui justifie la pertinence du présent article qui abordera les grandes tendances adoptées par les concepteurs de LAADQ, les pistes pour faciliter un choix judicieux et le type d'analyse habituellement réalisée avec ces logiciels. L'objectif de cet article n'est pas de juger de la pertinence de l'utilisation des logiciels en recherche qualitative, ce que de nombreux auteurs ont déjà fait (notamment Dumont, 2010; Hamel 2010), ni d'arriver à un classement consensuel des logiciels, ce qui serait hautement improbable. L'usage des LAADQ fait désormais partie des mœurs en recherche qualitative, qu'on le veuille ou non, et selon nous, leur usage est pertinent s'il découle d'un choix éclairé et d'une utilisation adéquate. Étant donné la grande activité dans le milieu, nous avons choisi d'examiner les logiciels qui ont récemment proposé une nouvelle version, avec une littérature relativement récente (après 2007) et qui jouissent d'une certaine reconnaissance dans les milieux francophones en éducation.

Une typologie en trois points

La typologie proposée dans cet article se base sur le degré d'intervention nécessaire de l'utilisateur pour réaliser une analyse. Contrairement à Lejeune (2010) qui propose une typologie en quatre axes, basée sur les fonctionnalités des logiciels, nous avons retenu un modèle à trois familles. Les axes suggérés par Lejeune (2010), c'est-à-dire « montrer, calculer, explorer et analyser » (p. 15) permettent une très bonne description des logiciels (simples ou élémentaires), mais dans le contexte des logiciels avancés, les outils répondent rapidement à ces quatre dimensions. Nous suggérons, pour notre part, un continuum en trois points : l'analyse automatique, semi-automatique et manuelle. L'analyse automatique est celle où le logiciel fait la plus grande partie de l'analyse avec un minimum d'interventions de la part du chercheur. Dans ce type d'analyse, il sera souvent nécessaire de passer plus de temps à préparer les données brutes et à interpréter les résultats. Ces logiciels permettent surtout de faire de l'analyse lexicométrique ou des statistiques textuelles : « [...] l'analyse lexicométrique vise à dégager les régularités, spécificités et corrélations des formes graphiques dans un corpus [...] »

(Bonnafous & Temmar, 2007, p. 13). Par conséquent, ils sont limités à l'analyse du texte sur lequel seul un reformatage aura été réalisé avant les analyses. Parmi ces logiciels, nous retenons Iramuteq (Ratinaud, 2012), Alceste (Image, 2012a) et Lexico (Lexico, 2012).

L'analyse semi-automatique demande plus d'interventions de la part des chercheurs. Les logiciels privilégiant ce type d'analyse permettent de faire une analyse lexicométrique, mais également de l'analyse par catégorisation sémantique (basée sur un dictionnaire, un thésaurus ou un regroupement de mots). Pour ce faire, le chercheur devra créer les relations entre certains thèmes et des groupes de mots, souvent choisis par rapport au cadre conceptuel, mais également en fonction du contenu des entrevues. Il est ensuite possible d'utiliser le travail réalisé dans un premier temps avec un corpus, pour ensuite faire l'analyse automatique d'un autre corpus. Il s'agit d'un avantage lorsqu'on poursuit un programme de recherches qui exigent des analyses semblables à l'intérieur d'une même étude (modèle à mesures répétées ou multisites par exemple) ou d'une étude à l'autre (évaluation périodique d'un programme). Les autres avantages de cette méthode sont de bénéficier d'une assistance pendant la classification et de systématiser par la suite les analyses (Sebastiani, 2002). Plusieurs auteurs rendent disponibles, et ce, parfois librement, des dictionnaires de codage permettant d'étudier des données à partir d'un cadre conceptuel bien précis (Hogenraad & Oriane, 1986; Péladeau, 2012a; Young & Soroka, 2011); ces dictionnaires peuvent aussi servir de base pour bâtir le sien. Les logiciels retenus dans cette catégorie sont Sato (Centre d'analyse de texte par ordinateur, 2012), Semato (Plante, 2012) et WordStat 6 (Péladeau, 2012b).

Enfin, l'analyse manuelle est celle où le chercheur doit le plus investir, mais aussi celle qui lui donne le plus de liberté et qui se rapproche le plus des méthodes traditionnelles d'analyse. Ces logiciels aident au codage des unités de sens, facilitent la classification des données et fournissent une assistance précieuse lors de l'analyse et de la gestion des rapports. Il est ainsi possible de générer des tableaux, des graphiques et des rapports en lien avec les données. La plupart des outils récents proposent également quelques solutions simples pour aider à faire du codage semi-automatique, mais rien d'aussi puissant que les logiciels mentionnés précédemment. Il est enfin possible de travailler avec des données qualitatives de type texte, image, son et vidéo. Les logiciels retenus dans cet axe sont : ATLAS.ti 7, (Scientific Software, 2012), HyperResearch (HyperResearch, 2012), QDA Miner 4 (Péladeau, 2012c), MAXQDA 11 (VERBI GmbH, 2012), NVivo 10 (QSR International, 2012) et Weft QDA (Weft QDA, 2012).

Il est important de rappeler que malgré une classification de chacun des outils dans un seul axe, la réalité est plus nuancée. Les dernières versions des LAADQ, et plus particulièrement des logiciels haut de gamme, proposent souvent des outils permettant une diversification des types d'analyse. Nous avons plutôt retenu l'utilité perçue, et souvent celle responsable de l'émergence du logiciel comme privilégiée.

Analyse automatique

Les logiciels d'aide à l'analyse automatique s'intéressent principalement à la structure lexicale du texte. Ils ne s'intéressent pas au sens des mots, mais plutôt aux mots, à la racine des mots et, dans certains cas, à des groupes de mots. Bien que ce type d'analyse soulève parfois la controverse dans le milieu de l'analyse de données qualitatives, nous la considérons comme un moyen pertinent pour plusieurs types de recherche tels que l'analyse de discours politique ou l'analyse de documents d'archives (Bonnafous & Temmar, 2007; Lamalle & Salem, 2002).

La première chose à savoir est que ces logiciels travaillent uniquement sur des documents texte, qui doivent être formatés pour répondre adéquatement aux exigences. Chaque logiciel possède ses propres spécificités et il est essentiel de consulter les guides d'instructions pour préparer les documents d'analyse. S'ils supportent plusieurs langues, le corpus doit être dans une seule langue, puisqu'il cherche à regrouper les formes lexicales qui se ressemblent.

Parmi les trois logiciels retenus, Iramuteq est un logiciel libre et ouvert. Il est assez récent, mais il a été retenu, car il semble prometteur et il est gratuit. L'analyse qu'il propose se base, entre autres, sur un système de classification très similaire à celui d'Alceste, un logiciel payant. Cette caractéristique de logiciel libre de droits vise toutefois à ce que l'analyste puisse faire preuve d'une plus grande autonomie, car le soutien que peut fournir une compagnie qui vend son logiciel n'est pas offert avec le logiciel ouvert. Alceste et Iramuteq proposent une analyse à partir de ce qu'ils appellent des « unités de contexte élémentaires » ou U.C.E. Bien que ces unités ne soient pas équivalentes à des phrases grammaticales, la manière de les diviser sous-tend plusieurs caractéristiques correspondant à la phrase². C'est à partir de ces U.C.E., composées d'un groupe d'environ 10 à 15 mots, que le logiciel détermine une corrélation entre les mots présents dans les textes. Ainsi, des mots qui se trouvent dans une même U.C.E., et ce, à plusieurs reprises, ont une corrélation plus forte entre eux. Le logiciel regroupe ensuite les U.C.E. semblables (les plus corrélées les unes avec les autres) sous forme de classe ou profil. Ces classes sont représentées par des mots clés (mots qui sont les plus fréquents dans les U.C.E.), par des formes représentatives (noms, prénoms,

adjectifs, etc.), ou encore par des catégories représentatives de la classe (genre, âge, statut socioéconomique, etc.). À noter que les logiciels passent préalablement par une phase de lemmatisation rendant ainsi possible de regrouper les mots d'une même famille (en fonction du genre et du nombre des noms ou encore du mode des verbes). Les logiciels permettent, entre autres, de réaliser une analyse factorielle de correspondance. La Figure 1 représente le résultat de l'analyse et permet de visualiser plusieurs éléments, comme les unités récurrentes (taille du texte), la proximité des termes (distance entre les mots) et les différentes classes (couleur des mots). Ce graphique permet d'obtenir relativement rapidement une compréhension du corpus. D'autres outils de visualisation aident à interpréter les relations entre les mots employés dans les différents corpus, toujours sur la base de la proximité des termes dans le texte.

Les logiciels Alceste et Iramuteq fournissent également un éventail de statistiques descriptives sur l'ensemble du corpus : le nombre et la fréquence des mots, le nombre d'U.C.E. classées, les formes lexicales, et beaucoup plus. C'est l'analyse et l'interprétation de ces classes qui permettent aux chercheurs de déterminer les grands thèmes récurrents d'un corpus et d'en mieux comprendre le contenu. Pour des analyses de base, et même intermédiaires, les deux logiciels fournissent les outils pertinents pour les données, mais Alceste en tant que logiciel payant, assure une meilleure offre de soutien, avec des formations et des guides complets (Image, 2012b). Il propose en outre un éventail plus grand d'outils de visualisation, tels que la cartographie en unités textuelles, la classification ascendante des formes, l'analyse factorielle en 3D ou encore la cartographie des résultats (Image, 2012a). La version 2012 permet maintenant une classification manuelle des unités d'analyse. Pour les utilisateurs avancés, il est préférable d'évaluer les outils que proposent l'un et l'autre des logiciels avant d'opérer un choix. À noter que l'évolution du logiciel Alceste est davantage garantie, considérant sa popularité et son coût. Le développement des logiciels libres est plus imprévisible, mais peut souvent aboutir à des projets réussis, comme en témoignent les nombreux logiciels libres du site Source Forge (www.sourceforge.net).

Pour sa part, Lexico se concentre plutôt sur les mots ou les segments répétitifs. Ainsi, pour un corpus donné, il fournira une liste des segments répétés avec ou sans leur contexte (Figure 2), une localisation des segments dans chaque texte, le décompte de chaque mot, etc. La localisation des mots est un ajout original très intéressant qui permet de visualiser l'organisation des idées dans un texte. Ainsi, Lexico permet de mettre en lumière l'utilisation du vocabulaire, son évolution d'un texte à l'autre ou encore les différences

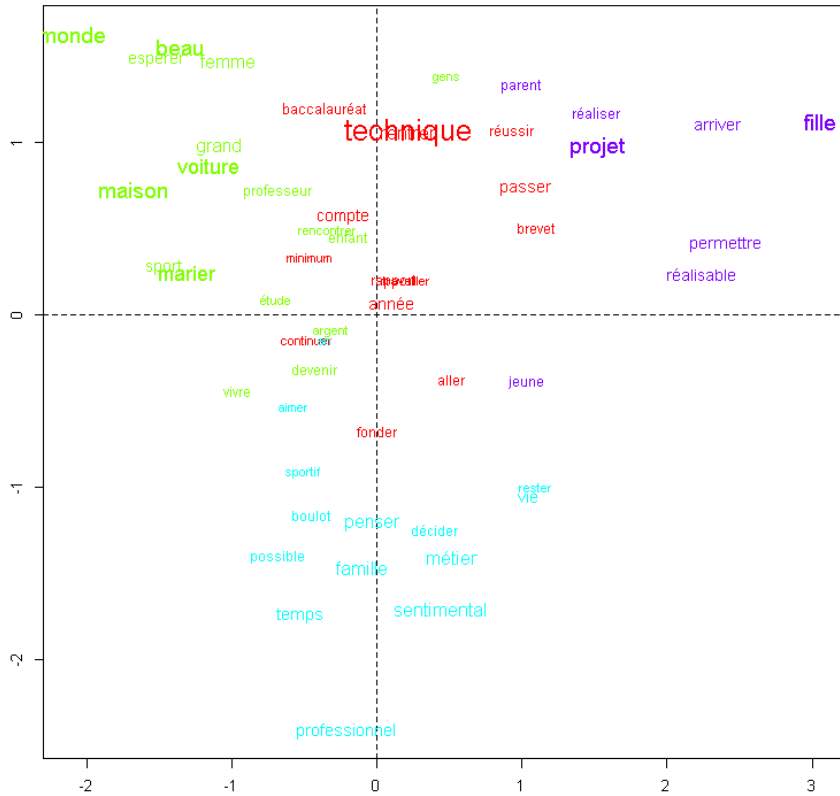


Figure 1. Analyse factorielle de correspondance (réalisée à partir du fichier de démonstration Alceste).

Maitre **Corbeau** , sur un arbre perché , Tenait
 é ! bonjour , Monsieur du **Corbeau** . Que vous êtes joli ! que
 es bois . " À ces mots le **Corbeau** ne se sent pas de joie ;
 omage , sans doute . " Le **Corbeau** , honteux et confus , Jura

Figure 2. Correspondance des formes.

stylistiques entre différents textes. Il est très utile pour des textes officiels (discours, documents d'archives, etc.). Pour des exemples concrets, le site de Lexico propose plusieurs résultats de recherche produits par les concepteurs du logiciel (Lexico, 2012).

L'une des plus grandes forces de l'analyse automatique, c'est la possibilité d'analyser un très large corpus beaucoup plus rapidement que ne le

ferait un traitement manuel. De plus, ces analyses assurent une très grande fiabilité, aux dépens toutefois d'une bonne validité. Selon Dorvil (2007), la fiabilité « fait référence à la reproductibilité de la mesure » (p. 396), alors que la validité « fait référence à la capacité d'un instrument de bien mesurer le phénomène à l'étude » (p. 396). La fiabilité d'une analyse n'est pas un gage de sa validité; le logiciel peut très bien répéter systématiquement la même erreur. Alors que, d'un point de vue lexical, l'unité « J'aime la tarte aux pommes » ressemble beaucoup à l'unité « Je n'aime pas la tarte aux pommes », le sens est complètement différent; or, une analyse automatique peut difficilement détecter ce genre de distinctions et les considérer à répétition comme faisant partie de la même classe.

Le traitement automatique comporte d'autres limites importantes. Comme il ne tient pas compte du sens de la phrase, mais uniquement du vocabulaire utilisé, la souris, mammifère, et la souris, outil informatique, y seront traitées comme un même mot. De plus, les analyses automatiques sont dépendantes de la langue, de la syntaxe et de la grammaire du texte. Pour certains corpus tirés de milieux « naturels » (forums en ligne, textes rédigés par des enfants, etc.) qui pourraient comporter des fautes ou des différences dans le langage utilisé, il pourrait alors être nécessaire d'utiliser un autre type d'analyse ou de réaliser un premier travail sur le corpus pour s'assurer de la qualité des analyses, à moins que l'objectif soit d'analyser les erreurs grammaticales. Enfin, il faut mentionner qu'une base en analyse quantitative, avec une compréhension des corrélations, des tests du chi-carré ou encore de l'analyse factorielle est la plupart du temps nécessaire pour bien interpréter les rapports produits par ces logiciels.

Analyse semi-automatique

Une deuxième gamme de logiciels permet une assistance considérable dans le travail d'analyse. Ces logiciels se basent en partie sur la structure lexicale du texte, mais laissent aux chercheurs la liberté de paramétrer et d'approfondir leur analyse. Ils rendent, entre autres, possible de travailler à partir de lexiques ou de thésaurus adaptés à son domaine ou à sa langue, ce qui apportera des nuances à l'analyse. Parmi les LAAQD, nous en avons retenu trois différents : Sato, Semato et WordStat.

Sato est un logiciel gratuit, disponible sur le site de l'Université du Québec à Montréal (<http://www.ling.uqam.ca/sato/>). Au départ, il fait un premier tri basé sur le découpage lexicologique, comme peuvent le faire les outils automatiques. Il est alors possible de traiter les différentes formes grammaticales du texte (adjectif, nom commun, pronom, etc.). Il permet toutefois d'aller plus loin en procédant à des regroupements de sens; le

chercheur utilise un thème donné (ex. : les animaux) et il crée un regroupement lexical autour de ce thème (ex. : *chat*, *chien*, *cheval*, *vache*, etc. réfèrent à des animaux.). La suite des analyses est réalisée sur ces nouveaux regroupements lexicaux. Les analyses se limitent principalement à des tableaux de statistiques, des tableaux croisés et des tableaux de fréquences. Les usages sont nombreux et variés, comme en témoigne la section « Les usages pratiques » du site du logiciel³. L'offre de formation est également présente et offerte par le Centre d'analyse de texte par ordinateur⁴.

Semato provient également du Centre d'analyse de texte par ordinateur de l'Université du Québec à Montréal et propose une plus grande gamme de services. Il existe deux facettes à cette plateforme : Sémato Sondage et Sémato Texte. La partie Sondage permet aux chercheurs de gérer et d'organiser un questionnaire en ligne, avec des questions ouvertes et fermées. Le module d'analyse qualitative est plutôt associé à la partie Texte. Semato n'est pas un logiciel à proprement parler, tout comme Sato d'ailleurs, car il n'y a pas d'installation à faire. L'analyse passe par un service Web⁵, ce qui facilite le partage et la collaboration avec des collègues. L'analyse semi-automatique se fait à partir de l'unité d'analyse préalablement choisie par l'utilisateur (réponse courte, chaque prise de parole en focus groupe, chaque paragraphe, etc.) que l'on doit identifier dans notre fichier d'analyse. Le système fournit une assistance dans la création d'une liste de vocables (termes autour d'un thème donné), ce qui s'apparente à un dictionnaire ou un thésaurus; ensuite, il classe automatiquement tous les contenus textuels. Cette classification peut, par la suite, être modifiée et adaptée selon les contenus et la compréhension du chercheur. À cela, le service Semato ajoute également une analyse automatique où il utilise comme unité de départ des synapsies. Elles représentent une courte unité contextuelle autour des mots clés récurrents. À partir des synapsies, Semato peut ensuite créer un champ sémantique autour de certains thèmes. Ainsi, il est possible de déterminer les mots les plus récurrents autour d'un autre terme. Il permet ainsi de créer une carte conceptuelle en fonction de la récurrence des mots (Figure 3). Pour visualiser et traiter ces classifications (champ sémantique ou thématization), Semato offre une vaste gamme d'outils graphiques et de tableaux.

Finalement, il y a WordStat, un module conçu pour fonctionner de manière intégrée aux logiciels QDA Miner, un logiciel d'analyse manuelle, ou SimStat, un logiciel d'analyse statistique. Cette caractéristique le place en bonne position pour les chercheurs tentés par les approches mixtes (Lewis & Maas, 2007). WordStat n'est en effet accessible que par l'un ou l'autre de ces logiciels pour amorcer l'analyse. Ainsi, à la suite d'une classification manuelle

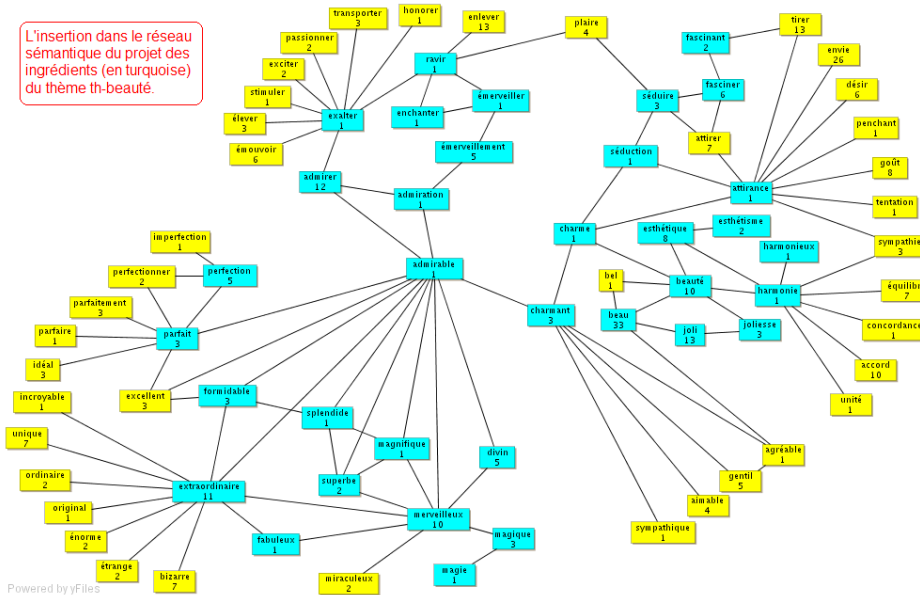


Figure 3. Réseau sémantique (tiré du site <http://semato.uqam.ca/guidexpertato/tuto5.asp>).

effectuée avec QDA Miner, il serait possible de faire une analyse automatique ou semi-automatique à l'aide de WordStat. Par exemple, ayant codé de façon manuelle des documents selon des grandes catégories comme Introduction, Objectifs, Moyens, Résultats et Conclusion, une analyse semi-automatique pourrait se concentrer uniquement sur le corpus codé Objectifs, fournissant ainsi la récurrence de certains thèmes sur ces parties spécifiques. Cette combinaison des analyses est une particularité de la suite de logiciels Provalis. D'abord et avant tout, WordStat offre de travailler avec des dictionnaires préexistants ou un dictionnaire personnel. Le dictionnaire (ou thésaurus dans certains logiciels) est composé de groupes de mots qui sont associés aux textes analysés. Ensuite, le logiciel propose toute une série de tableaux et graphiques qui permettent de travailler soit sur la thématisation réalisée ou à partir des mots présents (Figure 4).

Comme l'analyse automatique, l'analyse semi-automatique est très fiable. Toutefois, la validité dépend directement du travail des chercheurs : « En fin de compte, il est judicieux de rappeler que l'analyse assistée par ordinateur permet une parfaite fiabilité mais que sa validité dépend largement



Figure 4. Dendrogramme des thèmes en lien avec le thésaurus (réalisé à partir du fichier de démonstration QDA Miner).

de la connaissance du corpus de données de la part du chercheur »⁶ [traduction libre] (Pollach, 2010, p. 744). Il faut un travail très minutieux de la part des chercheurs lors des premières phases de travail, soit lors de la réalisation, mais surtout lors de la validation du dictionnaire. Il faut également souligner que, puisque c'est le chercheur qui contrôle le contenu du thésaurus, il devient possible d'analyser un corpus qui inclut plusieurs langues.

Bien qu'assez comparables dans l'interprétation des résultats, les méthodologies sous-jacentes aux logiciels ne le sont pas. Le choix des unités d'analyse, la terminologie utilisée, les algorithmes et les outils de visualisation sont assez différents et peuvent demander un certain temps d'adaptation. L'appropriation de ce type de logiciels est habituellement un peu moins intuitive que ne l'est celle des LAADQ utilisant une méthode manuelle.

Analyse manuelle

Les logiciels appartenant à la catégorie de l'analyse manuelle offrent surtout un soutien aux analyses. Ils permettent aux chercheurs d'être assistés pour la préparation des documents, le codage et la création de rapports. Le logiciel ne fait pas le travail de réflexion à la place du chercheur. Il fournit un espace structuré pour organiser ses idées. Il facilite le transfert d'idées entre chercheurs, et fournit un cadre plus formel. En ce sens, Basit (2003) conclut que :

[...] le codage par voie électronique rend le processus d'analyse moins lourd, mais il faut toutefois calculer le temps considérable

que l'on doit d'abord consacrer à s'initier au logiciel choisi. L'ordinateur facilite également l'analyse approfondie des données et les rapports générés sont très avantageux. Dans tous les cas, le codage reste toujours un exercice intellectuel⁷ [traduction libre] (p. 152).

Nous avons retenu six logiciels dans la catégorie de l'analyse manuelle, probablement les logiciels les plus connus dans le domaine : Weft QDA, HyperResearch 3 (seul logiciel disponible sur MAC), QDA Miner 4, ATLAS.ti 7, MAXQDA 11 et NVivo 10. Ces logiciels n'offrent pas la même gamme d'outils de traitement (fréquence de codes, recherche de codes, corrélations...), mais offrent tous les mêmes outils d'analyse de base (élaboration d'une grille de codes, codage de différentes façons, classification des documents, rapports de base, etc.). Par exemple, ils permettent tous de découper le texte en unités de sens et d'y accoler des étiquettes, codes ou thèmes. Nous insisterons davantage sur les spécificités de chaque logiciel.

Weft QDA est un logiciel libre et ouvert. Il possède peu d'options de production de rapports, mais permet de réaliser un découpage et des analyses de base : liste des unités de sens, tableaux de fréquence et tableaux croisés. Il se limite au traitement de fichier texte, sans aucun formatage (couleur, tableau, gras, italique, etc.). Pour le chercheur qui désire travailler à partir d'une interface simple pour ne pas avoir à investir trop de temps à l'apprentissage d'un logiciel, Weft QDA permet de facilement s'y retrouver, avec un guide d'utilisateur d'à peine une vingtaine de pages. Toutefois, le développement minimal au point de vue de la conception joue en sa défaveur aux plans de la convivialité et de l'interface d'utilisateur. Gratuit, il peut toutefois remplir les besoins d'étudiants en formation.

Pour aller plus loin dans l'analyse, HyperResearch possède davantage d'outils. Non seulement il permet de générer des rapports de résultats, mais il permet également de créer des combinaisons de codes complexes, de regrouper en catégories, de formuler des mémos ou même de représenter visuellement le codage. Il permet l'analyse de fichier de texte non formaté, de fichier audio ou de fichier vidéo. Il s'agit du seul logiciel disponible sur le système d'exploitation MacOSX. Toutefois, il possède une interface assez rudimentaire, ce qui peut rendre son utilisation un peu plus complexe, mais après quelque temps, l'utilisateur s'y retrouve assez facilement (pour voir l'interface : <http://www.researchware.com/>). Il est par contre moins polyvalent que certains de ces compétiteurs, que nous décrivons ci-dessous, en ne permettant qu'un nombre limité d'analyses et en offrant un outil de réseaux conceptuels plutôt rudimentaire (Figure 5).

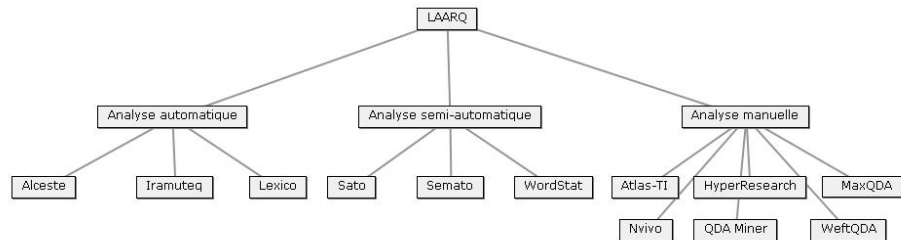


Figure 5. Exemple d'un réseau conceptuel élaboré avec HyperResearch.

QDA Miner est un logiciel assez récent dans le marché par rapport aux autres concurrents. Il supporte tous les types de document en format texte, mais également des PDF, des fichiers de bases de données ou encore des images. Il s'agit d'un atout important pour ce logiciel. Il permet très facilement l'analyse de questionnaires récoltés en ligne. Il supporte le traitement des données tant qualitatives que quantitatives, de même que les croisements entre celles-ci, utile dans les recherches avec des approches mixtes, comme mentionné précédemment. Il possède les outils de base (fréquence des codes, liste des unités de sens, occurrence dans chaque document, etc.), une bonne diversité d'outils d'analyse quantitative (corrélation de Pearson, F de Student, chi-carré, etc.) et un large éventail de graphiques (Figure 6). Cependant, la force de QDA Miner est également sa faiblesse. L'accent est mis sur l'analyse quantitative de données qualitatives et, par conséquent, les chercheurs plus proches des méthodes qualitatives traditionnelles s'y retrouvent un peu moins.

ATLAS.ti, MAXQDA et NVivo pourraient être considérés comme les logiciels les plus connus (et les plus anciens avec HyperResearch). Ils proposent un grand nombre d'outils d'analyse. Ainsi, ils permettent de coder à peu près tous les types de documents (textes, audio, vidéos, pages Web, PDF, etc.), possèdent des outils pour schématiser des réseaux conceptuels, génèrent des tableaux, des matrices et des rapports, etc.

Selon plusieurs, ATLAS.ti aurait d'abord été pensé pour soutenir l'analyse par théorisation ancrée (Larossa, 2005; Seale, Gobo, Gubrium, & Silverman, 2004). Sa terminologie, ses outils et son organisation laissent beaucoup de place aux chercheurs. Il fonctionne à partir d'un arbre de code à un seul niveau et facilite le codage émergent du texte. Il possède un outil avancé pour créer des réseaux conceptuels où il est possible d'ajouter des mémos ou des relations complexes entre les codes.

MAXQDA propose une interface très complète avec une multitude d'options : codage par couleur ou par émoticône, organisation hiérarchique des

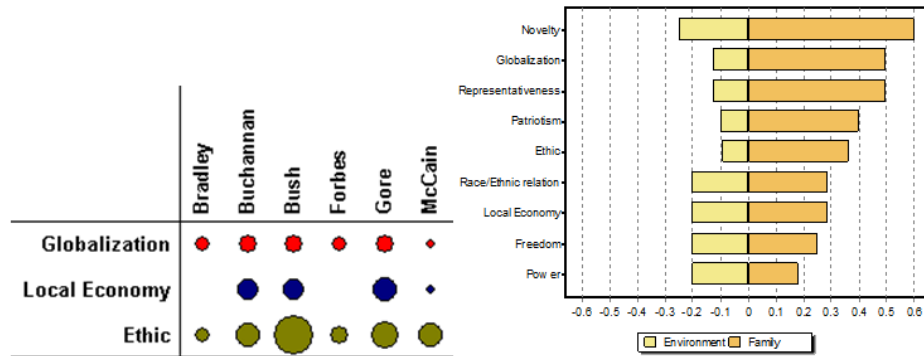


Figure 6. Exemples de graphiques à partir de QDA Miner. (réalisés à partir du fichier de démonstration QDA Miner).

documents et des codes, iconographie avancée dans le système de codage, ainsi qu'une multitude d'outils d'analyse (analyses des codes, méthodes mixtes, réseaux conceptuels ou MAXDictio⁸). Toutefois, cette vaste gamme d'outils rend l'apprentissage un peu plus complexe et ce n'est pas toutes les interfaces des outils qui sont faciles d'utilisation. Il y a d'abord une phase « d'activation »⁹ que l'on doit faire pour entreprendre toute analyse, ce qui demande une certaine éthique de travail. De plus, l'outil qui permet de générer les tableaux demande de construire une requête à partir de la logique booléenne (ex. : âge > 20 ou sexe = homme), ce qui n'est pas nécessairement facile au départ. Il n'en reste pas moins que certains chercheurs peuvent se passer complètement de ces outils et trouver leur compte dans ce logiciel, où l'on peut également faire des réseaux conceptuels. La plus récente version propose même une application pour tablette mobile, caractéristique unique par rapport aux autres logiciels.

NVivo génère davantage de tableaux et graphiques et utilise des outils qui rappellent davantage les méthodes quantitatives. Si ATLAS.ti fournit peu d'outils empruntés aux méthodes quantitatives, il permet d'exporter un fichier dans le logiciel IBM SPSS¹⁰; de son côté, NVivo permet de créer et de réaliser des réseaux conceptuels, malgré son penchant plus quantitatif. Bien que nous trouvions que NVivo fournit un plus grand support lorsque les données proviennent de sources variées et nombreuses (> 100), il est toutefois un peu plus difficile à apprendre au départ. En effet, il possède son langage propre et la diversité de ses outils d'analyse nuit à une prise en main rapide. Néanmoins, pour les projets complexes et de longue durée, il fournit une aide précieuse. À

noter que selon le site de QSR International, une version de NVivo pour MAC sera disponible en 2013.

Certains auteurs préféreront baser leur choix sur le type d'analyse. Il est possible de considérer deux sous-classifications de ces outils, bien que de moins en moins pertinentes considérant les nouvelles options des logiciels. L'analyse verticale (ou intracas) ou horizontale (intercas) se réfère au plan d'analyse des chercheurs (Miles & Huberman, 2003; Mukamurera, Lacourse, & Couturier, 2006). Pour l'analyse verticale, on cherche à décrire en profondeur un nombre limité de cas ou de situations. Dans la plupart des cas, les chercheurs tentent d'obtenir des données de plusieurs sources pour un même cas. L'analyse horizontale ressemble davantage à un modèle quantitatif. On cherche à comparer des variables et à établir des différences entre différents groupes de cas (homme et femme par exemple). QDA Miner, MAXQDA et NVivo sont considérés par certains chercheurs plus adéquats pour faire de l'analyse horizontale, alors qu'ATLAS.ti, HyperResearch et Weft QDA seraient plus appropriés pour l'analyse verticale.

Par ailleurs, tous ces logiciels proposent des versions gratuites plus légères qui permettent d'expérimenter et il s'agit souvent d'un bon moyen de familiarisation. Même si certains logiciels présentent des limites, il faut se rappeler que tous offrent les fonctionnalités de base et, en ce sens, répondent aux besoins d'une majorité de chercheurs. Puisqu'ils se rapprochent assez souvent des méthodologies plus traditionnelles, l'appropriation des LAADQ avec une approche manuelle se fait, en général, plus naturellement qu'avec les autres types de logiciels.

Conclusion

En guise de synthèse, le Tableau 1 brosse un tableau succinct de l'analyse proposée dans le présent article selon quatre critères, soit les formats de documents supportés par les logiciels, le degré d'expertise nécessaire pour les utiliser, la puissance d'analyse qu'ils permettent et leur coût. En appendice, nous vous proposons une liste non exhaustive d'articles assez récents, en français et en anglais, référant aux différents logiciels dans leur méthodologie. L'intérêt de cette liste n'est pas de juger de la pertinence ou même d'en faire une analyse critique, mais plutôt de présenter des applications pratiques des logiciels décrits dans le présent article. Sans juger de façon exhaustive de leur qualité, nous considérons que ces articles proposent une utilisation pertinente des outils d'analyse.

Le choix d'un logiciel n'est pas chose simple, mais leur utilisation n'est pas non plus une condition essentielle au succès d'un projet. Pour des projets de petites et moyennes envergures, le choix du chercheur peut très bien se faire

en fonction de la disponibilité du logiciel, de l'accessibilité à des formations, de la familiarité avec l'environnement numérique du logiciel, du coût d'acquisition ou, encore, de la langue du logiciel (par exemple, ATLAS.ti est disponible uniquement en anglais). Les projets de plus grandes envergures nécessitent de mieux cibler les besoins afin de s'assurer que le logiciel permettra d'y répondre de la manière la plus complète possible (documents à coder, types d'analyse anticipée, etc.).

Le travail en équipe est surtout favorisé dans les logiciels manuels, où le traitement d'analyse est beaucoup plus long. En ce sens, tous les outils, sauf Weft QDA, offrent des outils facilitant le travail en équipe. Il est possible de s'identifier en tant que codeur, d'importer et d'exporter entre les projets, ou dans certains cas, de calculer l'accord interjuges (pour NVivo, MAXQDA et QDA Miner). NVivo est le seul à offrir une version « serveur » qui permet à plusieurs codeurs de travailler de façon synchrone sur le même projet, sans avoir à faire plusieurs copies du projet.

L'analyse des LAADQ proposée ici représente bien humblement la pointe de l'iceberg. Les fonctionnalités présentées pour chaque logiciel ne dépassent souvent qu'une partie des options disponibles et nous avons présenté seulement quelques logiciels parmi la gamme des logiciels spécialisés. Que ce soit pour les analyses automatique¹¹, semi-automatique ou manuelle¹², il existe plusieurs ressources inexplorées. Force est de constater que les logiciels continuent d'évoluer. Seulement dans la dernière année, plusieurs nouvelles versions sont apparues sur le marché : Alceste 2012, ATLAS.ti 7, NVivo 10 et QDA Miner 4. Par conséquent, avant de sélectionner un logiciel, nous conseillons d'abord de cibler le modèle d'analyse qui vous intéresse, d'explorer les outils disponibles et d'évaluer les différentes publications réalisées à partir du logiciel. Pour les projets complexes (plusieurs types de documents, nombreuses collectes de données, multiples sites de collectes, etc.), il est conseillé de bien connaître les logiciels (soit à partir d'une formation ou d'une lecture exhaustive des documents d'aide) et de réfléchir à l'organisation des documents. Il faut également noter que l'utilisation de ces logiciels n'est pas mutuellement exclusive (Bonneau, 2008; Daoust, Dobrowolski, Dufresne, & Gélinas-Chebat, 2006; Duchastel, Daoust, & Della Faille, 2008; Fallery & Rodhain, 2007; Franzosi, Doyle, McClelland, Putnam Rankin, & Vicari, 2012).

L'avenir des logiciels semble se diriger vers l'intégration des outils d'analyse et la méthodologie mixte, comme le soulève Bazeley (2010) : « Les développements récents des LAADQ reflètent l'intérêt grandissant pour l'intégration des données qualitatives et quantitatives et de leurs analyses

Tableau 1
Synthèse des logiciels d'aide à l'analyse de données qualitatives

	Formats supportés¹³	Degré d'expertise suggéré¹⁴	Puissance d'analyse¹⁵	Coûts¹⁶ (tarif éducation, licence simple)
Alceste	Textes bruts	●●●	●●●	Sur demande
Iramuteq	Textes bruts	●●●●	●●●	Gratuit
Lexico 3	Textes bruts	●●	●●	150 €
Sato	Textes bruts	●●●	●●	Gratuit
Semato	Textes bruts	●●●●	●●●	100 CAD
WordStat 6	Textes bruts (+ Images codées avec QDA Miner)	●●●	●●●	590 US
ATLAS.ti 7	Textes avancés, images, sons, vidéos, pages Web	●●●	●●●	670 US (99 US étudiant)
HyperResearch	Textes bruts, images, sons, vidéos	●●●	●●	499 US (199 US étudiant)
NVivo 10	Textes avancés, images, sons, vidéos, pages Web	●●●●	●●●●	670 US (215 US étudiant)
MAXQDA 11	Textes avancés, images, sons, vidéos, pages Web	●●●●	●●●●	560 US (99 US étudiant)
QDA Miner 4	Textes avancés, images, pages Web	●●	●●●●	590 US
Weft QDA	Textes bruts	●	●	Gratuit

comme une approche légitime dans le traitement de données »¹⁷ [traduction libre] (p. 434). En ce sens, les chercheurs auront à réfléchir sur les usages des différents outils à leur disposition afin de valoriser de bonnes pratiques de recherche.

Notes

¹ Rappel : logiciels d'aide à l'analyse de données qualitatives.

² Dans l'analyse plus traditionnelle, on réfère plutôt aux unités de sens (Van der Maren, 1996), aux unités de contexte (Bardin, 1977) ou encore aux unités d'analyse (Miles & Huberman, 2003), qui peuvent référer à un mot, une phrase, un paragraphe, une idée, une image, etc. Comme on peut le constater, dans l'analyse traditionnelle, la phrase n'est qu'un choix d'unité parmi d'autres.

³ http://www.ling.uqam.ca/sato/satoman-fr_presentation.html#usages

⁴ <http://www.ling.uqam.ca/ato/>

⁵ « Composant applicatif accessible sur le Web, par l'entremise d'une interface standard, qui peut interagir dynamiquement avec d'autres applications en utilisant des protocoles de communication basés sur le XML, et cela indépendamment du système d'exploitation et des langages de programmation utilisés » (OQLF, 2004).

⁶ « *Ultimately, one has to bear in mind that computer-assisted text analysis has perfect reliability, but its validity depends hugely on how well the researcher knows the data set* » (Pollach, 2010, p.744).

⁷ « [...] *electronic coding made the process relatively smooth, though considerable time had to be spent initially to get acquainted with the package. The computer also facilitated the analyses to be carried out in more depth and the reports generated were invaluable. Nevertheless, coding was an intellectual exercise in both the cases* » (Basis, 2003, p. 152)

⁸ Analyse lexicologique des contenus.

⁹ L'activation indique au logiciel les éléments (documents et codes) sur lesquels on veut faire des analyses.

¹⁰ IBM SPSS est un logiciel pour l'analyse de données qualitatives. L'exportation d'un projet de recherche qualitatif vers un fichier quantitatif passe par une conversion dichotomique (0 ou 1) du codage. Chaque unité de sens devient une observation et chaque code devient une variable. Un code présent équivaut à 1 et un code absent donne un 0.

¹¹ <http://www.kdnuggets.com/software/text.html>

¹² <http://www.eval.org/Resources/QDA.asp>

¹³ La mention « Textes bruts » fait référence aux formats de texte sans mise en page avancée (.txt, .rtf) alors que « Textes avancés » indique plutôt des fichiers plus avancés (.doc, .docx, .pdf, .opd, etc.).

¹⁴ Le degré d'expertise nécessaire pour l'utilisation est basé sur la complexité du langage, de même que la facilité de navigation et des différentes options disponibles, selon notre perception (● = Facile d'utilisation; ●●●● = Complexe)

¹⁵ La puissance d'analyse suggère la diversité des analyses offertes (rapports texte, tableaux, graphiques, outils de visualisation) et des outils (● = Peu d'outils d'analyse; ●●●● = Nombreuses possibilités d'analyse et d'exportation)

¹⁶ Tirés des sites respectifs, en décembre 2012, avant taxes. D'autres versions sont disponibles (licence de 1 an, pour compagnies, en réseau, etc.).

¹⁷ « *Recent developments in software reflect the growing in acceptance of the integration of qualitative and quantitative data and analyses as a legitimate approach to handling data* » (Bazeley, 2010, p. 434).

Références

- Al Kadri, H. M. F., Al-Moamary, M. S., Elzubair, M., Magzoub, M. E., AlMutairi, A., Roberts, C., & van der Vleuten, C. (2011). Exploring factors affecting undergraduate medical students' study strategies in the clinical years : a qualitative study. *Advances in Health Sciences Education, 16*(5), 553-567.
- Ali, K. (2009). Changements des politiques d'immigration et d'asile après les attentats du 11 septembre 2011 : rupture ou continuité? (Mémoire de maîtrise inédit). Université de Montréal, QC.
- Audet, R. (2011). Le Sud global et les nouvelles figures de l'équité à l'OMC. *Éthique et économique, 8*(2), 30-47.
- Bardin, L. (1977). *L'analyse de contenu*. France : Presses universitaires de France.
- Barkley, C. (2012). *School leader use of social media for professional discourse*. Virginia : Virginia Commonwealth University.
- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research, 45*(2), 143-154.
- Bazeley, P. (2010). Computer-assisted integration of mixed methods data and sources analyses. Dans A. Tashakkori, & C. B. Teddlie (Éds), *Sage handbook of mixed methods in social and behavioral research* (2^e éd., pp. 431-467). Thousand Oaks, CA : Sage.
- Beaupré-Laforest, C. (2007). *Les transformations du discours de la gauche au Québec à travers le passage du providentialisme au néolibéralisme* (Mémoire de maîtrise inédit). Université de Montréal, QC.

- Bonnafous, S., & Temmar, M. (2007). *Analyse du discours et sciences humaines et sociales*. Paris : Éditions Ophrys.
- Bonneau, J. (2008). Outils d'aide à l'exploitation d'entretiens semi-directifs : étude de l'interaction entre intervieweur et interviewés sur un corpus ethnoécologique. Dans S. Heiden, & B. Pincemin (Éds), *Actes des 9^{es} journées d'analyse de données textuelles* (pp. 219-231). Lyon : Presses universitaires de Lyon.
- Boritz, J. E., Hayes, L., & Lim, J.- H. (2012). A content analysis of auditors' reports on IT internal control weaknesses : the comparative advantages of an automated approach to control weakness identification. *International Journal of Accounting Information Systems*. Repéré à <http://www.sciencedirect.com/science/article/pii/S1467089511000674>
- Bourion, C., & Persson, S. (2011). La transmission intergénérationnelle des talents par le mentoring épistolaire. *Revue internationale de psychosociologie*, 17(41), 321-346.
- Caron, J., & Portelance, L. (2012). Appropriation et intégration des changements en éducation par des enseignants associés dans leurs pratiques de formation des stagiaires. *Éducation et francophonie*, 40(1), 176.
- Centre d'analyse de texte par ordinateur. (2012). *Système d'analyse de texte par ordinateur*. Repéré à <http://www.ling.uqam.ca/sato/>
- Clark, J. A. (2010). *Exploration of children's literature core-curriculum alignment with preservice teacher practicum experience*. ProQuest Dissertations and Theses. Northcentral University, Arizona.
- Collin, S., Karsenti, T., & Dumouchel, G. (2011). *L'interaction en ligne pour développer la pensée réflexive des futurs enseignants : patterns réflexifs et influence de l'outil*. Repéré à <http://edutice.archives-ouvertes.fr/edutice-00676174/>
- Coulon, A. (1992). *L'École de Chicago*. Paris : Presses universitaires de France.
- Courtinat-Camps, A., de Leonardis, M., & Preteur, Y. (2011). Formes du rapport au savoir chez des collégien(ne)s à haut potentiel intellectuel. *Neuropsychiatrie de l'enfance et de l'adolescence*, 59(6), 336-341.
- Dalud-Vincent, M. (2010). Les "choix" du sociologue avec Alceste. Du paramétrage des unités de contexte aux résultats. *BMS. Bulletin de méthodologie sociologique*, 107, 23-48.

- Dalud-Vincent, M. (2011). Alceste comme outil de traitement d'entretiens semi-directifs : essai et critiques pour un usage en sociologie. *Langage et société*, 135, 9-28.
- Daoust, F., Dobrowolski, G., Dufresne, M., & Gélinas-Chebat, C. (2006). Analyse exploratoire d'entrevues de groupe : quand ALCESTE, DTM, LEXICO et SATO se donnent la main. *Les Cahiers de la MSH Ledoux*, 1(3), 313-326.
- Donot, M. (2010). Émergence d'un nouveau péronisme? Analyse des discours à la Nation de Néstor Kirchner (2003-2007). *Mots. Les langages du politique*, 93. Repéré à <http://www.cairn.info/revue-mots-2010-2-p-65.htm>
- Dorvil, H. (2007). *Problèmes sociaux. Tome III. Théories et méthodologies de la recherche*. Québec : Presses de l'Université du Québec.
- Duchastel, J., Daoust, F., & Della Faille, D. (2008). Le problème de l'interprétation des données à partir d'un corpus bilingue. L'exemple du discours des trois chefs de la partie sur la motion de reconnaissance du « Québec comme société distincte au sein du Canada ». Dans S. Heiden, & B. Pincemin (Éds), *Actes des 9^{es} journées d'analyse de données textuelles* (pp. 412-431). Lyon : Presses universitaires de Lyon.
- Dufresne, M., Dupuis, F., & Tremblay, M. (2008). *La préverbation en français médiéval : polysémie et sens grammatical*. Actes du Congrès Mondial de Linguistique Française (CMLF). Repéré à http://www.linguistiquefrancaise.org/index.php?option=com_article&access=doi&doi=10.1051/cmlf08280&Itemid=129
- Dumont, V. (2010). Introduction. Du débat sur la place des logiciels dans l'analyse de données qualitatives. *Recherches qualitatives, Hors-série*, 9, 1-14.
- Durand, M.-J., & Poirier, S. (Éd.). (2012). La recherche collaborative au service du développement professionnel en ligne. *Éducation et francophonie*, XL(1).
- Estryn-Behar, M. (2010). Comment des médecins hospitaliers apprécient leurs conditions de travail. Réponses aux questions ouvertes d'une enquête nationale. *Revue française des affaires sociales*, 4(4), 27-52.
- Éthier, M.-A., & Lefrançois, D. (2007). *Le transfert par des élèves du primaire des compétences développées en classe de sciences sociales*. Actes des Journées d'étude didactique de l'INRP. Lyon.

- Fallery, B., & Rodhain, F. (2007, Juin). *Quatre approches pour l'analyse de données textuelles : lexicale, linguistique, cognitive, thématique*. Actes de la XVI^e Conférence internationale de management stratégique. Montréal, QC. Repéré à http://www.aims2007.uqam.ca/actes-de-la-conference/communications/falleryb300/at_download/article.pdf
- Franzosi, R., Doyle, S., McClelland, L., Putnam Rankin, C., & Vicari, S. (2012). Quantitative narrative analysis software options compared : PC-ACE and CAQDAS (ATLAS.ti, MAXqda, and NVivo). *Quality & Quantity*, 1-29.
- Gervais, C., & Leroux, M. (2011). Ressources mobilisées par des stagiaires pour enseigner : faible part des savoirs théoriques. *Revue des sciences de l'éducation*, 37, 281-306.
- Hamel, J. (2010). L'objet d'analyse comme pivot de l'analyse qualitative assistée par ordinateur. *Recherches qualitatives, Hors-série*, 9, 170-180.
- Hawi, N. (2010). Causal attributions of success and failure made by undergraduate students in an introductory-level computer programming course. *Computers & Education*, 54(4), 1127-1136.
- Hogenraad, R., & Oriane, E. (1986). Imagery, regressive thinking, and verbal performance in internal monologue. *Imagination, Cognition, and Personality*, 5(2), 127-145.
- HyperResearch. (2012). *ResearchWare*. Repéré à <http://www.researchware.com/products/hyperresearch.html>
- Image. (2012a). *Ingénierie et statistique textuelle*. Repéré à http://www.image-zafar.com/index_alceste.htm
- Image. (2012b). *Formations Alceste*. Repéré à <http://www.image-zafar.com/fr/formations-alceste-fr>
- Johnson, C. M., Sharkey, J. R., Dean, W. R., McIntosh, W. A., & Kubena, K. S. (2011). It's who I am and what we eat. Mothers' food-related identities in family food choice. *Appetite (Print)*, 57(1), 220-228.
- Lac, M. (2010, Septembre). *L'évolution des politiques éducatives et des représentations de l'éducation en France durant le 21^e siècle à travers l'analyse des « Rapports d'activités du ministère de l'éducation »*. Actes du congrès de l'Actualité de la recherche en éducation et en formation. Genève.

- Lamalle, C., & Salem, A. (2002). Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels. Dans A. Morin, & P. Sébillot (Éds), *Actes des 6^{es} journées d'analyse de données textuelles* (pp. 403-412). Saint-Malo : INRIA.
- Larossa, R. (2005). Grounded theory methods and qualitative family research. *Journal of Marriage and Family*, 67, 837-857.
- Laurens, E., Gimenez, G., & Barthelemy, S. (2011). Internet : de l'usage à l'addiction : analyse textuelle d'entretiens cliniques. *Alcoolologie et addictologie*, 33(3), 227-238.
- Lejeune, C. (2010). Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative. *Recherches qualitatives, Hors-série*, 9, 15-32.
- Lewis, R. B., & Maas, S. M. (2007). QDA Miner 2.0 : mixed-model qualitative data analysis software. *Field methods*, 19(1), 87-108.
- Lexico. (2012). *Lexico 3*. Repéré à <http://www.tal.univ-paris3.fr/lexico/>
- Lilley, S. C. (2008). Information barriers and Maori secondary school students. *Information Research : an international electronic journal*, 13(4), 331-344.
- Lussier, N., Bilodeau, A., & Bélanger, J. (2007). *Description et analyse de la pratique d'accompagnement scolaire dans deux services de soutien à la réussite scolaire du quartier Saint-Michel*. Montréal : Agence de la santé et des services sociaux de Montréal. Bibliothèque et Archives nationales du Québec.
- Marchand, P., & Ratinaud, P. (n.d.). Psychologie sociale de la communication : médias et politique. Les mots du "grand débat sur l'identité nationale". Repéré à <http://pascal-marchand.fr/spip.php?article42>
- Marty, E., Smyrniotis, N., & Rebillard, F. (2011). A multifaceted study of online news diversity : issues and methods. Dans R. Salaverria (Éd.), *Diversity of journalism, conference proceedings (CICOM/ECREA)* (pp. 228-242). Pamplona : Université de Navarre.
- Miles, M. B., & Huberman, A. M. (2003). *Analyse des données qualitatives* (2^e éd.). Paris : De Bœck.
- Mukamurera, J., Lacourse, F., & Couturier, Y. (2006). Des avancées en analyse qualitative : pour une transparence et une systématisation des pratiques. *Recherches qualitatives*, 26(1), 110-138.

- Mysore, N., Pluye, P., Grad, R. M., & Johnson-Lafleur, J. (2009). Tensions associated with the use of electronic knowledge resources within clinical decision-making processes : a multiple case study. *International Journal of Medical Informatics*, 78(5), 321-329.
- O'Halloran, K. (2011). Investigating argumentation in reading groups : combining manual qualitative coding and automated corpus analysis tools. *Applied linguistics*, 32(2), 172-196.
- OQLF. (2004). *Le grand dictionnaire terminologique*. Office québécois de la langue française. Repéré à http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8358885
- Péladeau, N. (2012a). *WordStat content analysis dictionaries*. Repéré à <http://www.provalisresearch.com/wordstat/Content-Analysis-Dictionaries.html>
- Péladeau, N. (2012b). *Wordstat v6.1*. Repéré à <http://www.provalisresearch.com/wordstat/Wordstat.html>
- Péladeau, N. (2012c). *QDA Miner v4*. Repéré à <http://www.provalisresearch.com/QDAMiner/Qualitative-Software.html>
- Plante, P. (2012). *Sémato Texte*. Repéré à <http://semato.uqam.ca/guidexpert-ato/ss.asp>
- Pollach, I. (2010). *Software Review : WordStat 5.0. Organizational Research Methods*.
- Primack, B. A., Nuzzo, E., Rice, K. R., & Sargent, J. D. (2012). Alcohol brand appearances in US popular music. *Addiction*, 107(3), 557-566.
- QSR International. (2012). *Introducing NVivo 10*. Repéré à <http://www.qsrinternational.com/>
- Raby, C., Karsenti, T., Meunier, H., & Villeneuve, S. (2011). Usage des TIC en pédagogie universitaire : point de vue des étudiants. *Revue internationale des technologies en pédagogie universitaire*, 8(3), 6-19.
- Ratinaud, P. (2012). *Iramuteq*. Repéré à <http://www.iramuteq.org/>
- ResearchWare. (2012). *About ResearchWare*. Repéré à <http://www.researchware.com/company/about-researchware.html>
- Sarai, N., & Sithole, G. (2012). Analysis of the process and outcomes of eportfolio development to recognise formal learning in Zimbabwe. *Computer and Information Science*, 5(3), 56-66.

- Scelles, R., Aubert Godard, A., Gargiulo, M., Avant, M., & Gortais, J. (2010). Revealing a child's pathology : physicians' experiences. *Early Child Development and Care*, 180(6), 743-752.
- Scientific Software (2012). About us. Repéré à <http://www.atlasti.com/aboutus.html>
- Seale, C., Gobo, G., Gubrium, J. F., & Silverman, D. (Éds). (2004). *Qualitative research practice*. London : Sage.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- Slim, M. (2007). *The parenting styles and the feeding practices of parents of diabetic adolescents* (Mémoire de maîtrise inédit). Université de Montréal, QC.
- Smyrnaiois, N., Marty, E., & Rebillard, F. (2010). Does the long tail apply to online news? A quantitative study of french-speaking news websites. *New Media & Society*, 12(8), 1244-1261.
- Truc, G. (2011). Analyser un corpus illisible? Le logiciel Alceste confronté à des registres de condoléances. *Langage et société*, 135(1), 29.
- Vaillancourt, S. (2011). *Les enseignants et la mise en œuvre du renouveau pédagogique au secondaire : une perspective locale* (Mémoire de maîtrise inédit). Université du Québec à Montréal, QC.
- Van der Maren, J. M. (1996). *Méthodes de recherche pour l'éducation*. (2^e éd.). Bruxelles : De Bœck.
- VERBI GmbH. (2012). *MaxQDA. The art of text analysis*. Repéré à <http://www.maxqda.com/>
- Villatte, A., Hugon, M., & de Leonardis, M. (2011). Forms of self-concept in gifted high school students enrolled in heterogeneous classes. *European Journal of Psychology of Education*, 26(3), 373-392.
- Weft QDA. (2012). *Weft QDA, A free qualitative analysis software application*. Repéré à <http://www.pressure.to/qda/>
- Young, L., & Soroka, S. (2011). Affective news : the automated coding of sentiment in political texts. *Political Communication*, 29(2), 205-231.
- Zakaria, N., Cogburn, D. L., Khadapkar, P. S., & Louis, C. (2012). Examining cultural effects on distributed decision-making processes using keyword analysis and data mining techniques. *International Journal of Business and Systems Research*, 6(3), 313-335.

Normand Roy est professeur à l'Université du Québec à Trois-Rivières au Département des sciences de l'éducation. Il s'intéresse tout particulièrement à l'utilisation des technologies en éducation. Il a passé les dix dernières années à faire de la recherche en éducation.

Roseline Garon est professeure agrégée à l'Université de Montréal depuis plus de 10 ans à la Faculté des sciences de l'éducation, Département d'administration et fondements de l'éducation. Ses recherches portent sur les pratiques des directions d'établissements et ses prestations de cours concernent principalement les méthodes de recherche, dont le cours « Analyse et interprétation des données qualitatives ».

Appendice

Nous proposons ici une sélection non exhaustive d'articles dans lesquels les auteurs, provenant de différents domaines, ont utilisé l'un ou l'autre des logiciels analysés. Les articles sont regroupés d'abord en fonction du logiciel utilisé, puis, selon le domaine abordé.

Tableau 2
Liste des références pertinentes en lien avec les logiciels d'analyse

	Domaine	Références pertinentes
Alceste	Éducation	Villatte, Hugon et de Leonardis (2011); Éthier et Lefrançois (2007)
	Psychologie	Bourion et Persson (2011); Courtinat-Camps, de Leonardis et Preteur (2011); Scelles, Aubert Godard, Gargiulo, Avant, & Gortais (2010)
	Sociologie	Dalud-Vincent (2010, 2011); Laurens, Gimenez et Barthelemy (2011); Truc (2011)
Iramuteq	Éducation	Lac (2010)
	Sociologie	Marchand et Ratinaud, (n.d.), Marty, Smyrnaiois et Rebillard (2011)
Lexico 3	Politique	Ali (2009); Donot (2010); Smyrnaiois, Marty et Rebillard (2010)
	Sociologie	Bonneau (2008); Estryn-Behar (2010)
Sato	Économie	Audet (2011)
	Littérature	Dufresne, Dupuis et Tremblay (2008)
	Sociologie	Beaupré-Laforest (2007)
Semato	Éducation	Lussier, Bilodeau et Bélanger (2007), Vaillancourt (2011)
	Médecine	Slim (2007); Mysore, Pluye, Grad et Johnson-Lafleur (2009)
WordStat 6	Variés	http://www.provalisresearch.com/Documents/SomeStudies.pdf

Tableau 2
Liste des références pertinentes en lien avec les logiciels d'analyse (suite)

	Domaine	Références pertinentes
ATLAS.ti 7	Éducation	Al Kadri, Al-Moamary, Elzubair, Magzoub, AlMutairi, Roberts et van der Vleuten (2011); O'halloran (2011)
	Nutrition	Johnson, Sharkey, Dean, Mcintosh et Kubena (2011)
	Sociologie	Primack, Nuzzo, Rice et Sargent (2012)
HyperResearch	Éducation	Hawi (2010); Lilley (2008); Raby, Karsenti, Meunier et Villeneuve (2011)
NVivo 10	Variés	http://www.qsrinternational.com/support_resource-articles.aspx?productid=0&categoryid=3
MAXQDA 11	Variés	http://www.maxqda.com/download/Literature_eng.pdf
QDA Miner 4	Comptabilité	Boritz, Hayes et Lim (2012)
	Éducation	Collin, Karsenti et Dumouchel (2011); Durand et Poirier (2012); Gervais et Leroux (2011)
	Sociologie	Zakaria, Cogburn, Khadapkar et Louis (2012)
Weft QDA	Éducation	Barkley, (2012); Caron et Portelance (2012); Clark (2010); Sarai et Sithole (2012)