

Corpus d'entretiens et analyse de données textuelles : un cas sur le changement organisationnel

Emmanuel Mastio¹, Jean Moscarola², Caroline Mothe³

¹ Professeur, University of Technology of Sydney

² Professeur, émérite, IREGE, Université Savoie Mont Blanc

³ Professeur des Universités, IREGE, Université Savoie Mont Blanc

Abstract

The automation of thesauri and linguistic engineering make textual data analyses evolve from lexical analysis to semantic analysis by giving it the capacity to restore categories of meaning which are standard or defined by the analyst. The exploration without a priori is thus enriched by the possibility of a targeted research in the manner of content analysis and coding. This particular approach to corpuses is an opportunity for an experience that is amplified by the graphic interaction and sharing possibilities of the web. This is what we want to show through the analysis thanks to the Sphinx iQ2 software of a corpus of 40 interviews collected during a research on the organizational change of a professional services firm specialized in intellectual property in Australia.

Keywords: Case study, lexical analysis, semantic analysis, Sphinx iQ2, textual data analysis

Résumé

L'analyse sémantique et l'ingénierie linguistique font évoluer l'analyse de données textuelles (ADT) de l'analyse lexicale à l'analyse sémantique en lui donnant la capacité de restituer des catégories de sens standard ou définies par l'analyste. L'exploration sans à priori se trouve ainsi enrichie par la possibilité d'une recherche ciblée à la manière de l'analyse de contenu et de la codification. Cette approche particulière des corpus est l'occasion d'une expérience amplifiée par les possibilités graphiques d'interaction et de partage du web. C'est ce que nous voulons montrer par l'analyse grâce au logiciel Sphinx iQ2 et DataViv d'un corpus de 40 entretiens recueillis à l'occasion d'une recherche sur le changement organisationnel d'un cabinet de services professionnels spécialisé en propriété intellectuelle en Australie.

Mots clés : Analyse de Données Textuelles, analyse lexicale, analyse sémantique, étude de cas, Sphinx iQ2

1. Introduction

L'expression « recherche qualitative » s'applique aux méthodes basées sur le recueil d'informations documentaires, la réalisation d'entretiens ou l'observation participante. Ces méthodes consistent à prendre connaissance des matériaux ainsi rassemblés pour proposer une interprétation argumentée par des citations de terrain et des références aux théories. Plus ou moins formalisé, ce travail d'analyse s'inscrit dans la tradition herméneutique et exclue le plus souvent le recours aux méthodes quantitatives (Dumez, 2013 ; Paillé, 2006).

L'analyse de données textuelles (ADT) se développe dans la deuxième moitié du XX^{ème} siècle au confluent des études littéraires, de la statistique, de la linguistique et de l'informatique (Benzecri, 1975 ; Lebart et al., 2020 ; Veronis, 2004). Cependant elle est peu utilisée par les chercheurs en sciences humaine tant les oppositions entre approche qualitative et quantitative restent prégnantes.

L'objectif de ce papier est de montrer comment les méthodes de l'ADT peuvent contribuer à l'analyse des corpus d'entretiens qui présentent un double défi pour le chercheur. Celui de la charge de travail quand il faut lire des corpus très volumineux et celui de l'objectivité quand il s'agit de répondre aux contraintes d'un travail scientifique.

Dans une première partie on exposera les enjeux de l'analyse de corpus d'entretiens et les principes et méthodes de l'ADT. Tout ceci sera illustré dans une seconde partie avec un corpus de 40 interviews portant sur l'étude d'un changement organisationnel. L'analyse menée avec les logiciels Sphinx et DataViv sera illustrée par des liens donnant la « vision ADT » de ce corpus. En cliquant sur ces liens lecteur pourra ainsi se faire sa propre opinion sur ces usages dont la portée et les limites seront discutées en conclusion.

2. Entretiens et analyse de données textuelles

2.1. Les méthodes traditionnelles d'analyse de corpus d'entretiens non directifs

Le recours aux entretiens non directifs, individuels ou de groupe est une pratique fréquente en science sociale. Elle se distingue des méthodes de collectes par voie de questionnaire par les objectifs poursuivis et les méthodes d'analyse mises en œuvre : Exploration versus confirmation, ambition de construire des théories versus volonté d'établir des connaissances testées et vérifiées, analyse qualitative versus analyse quantitative. A l'origine d'une recherche par entretiens, le chercheur se méfie de ses a priori et, s'il est guidé par un modèle, il cherche plus à le préciser ou à le faire évoluer qu'à le vérifier d'une manière systématique. La mobilisation des personnes interviewées répond à la volonté de refléter la variété des situations possibles plutôt que de représenter fidèlement une population comme on cherche à le faire en administrant un questionnaire.

2.1.1. Recherche qualitative pure

L'analyse de ces corpus renvoie à la tradition des recherches littéraires ou historiques consistant à lire, comprendre et interpréter le matériau rassemblé pour en rendre compte de manière descriptive, explicative, ou compréhensive. L'analyste, comme l'historien ou le journaliste, construit ainsi l'histoire qu'il écrit. Elle est autant le reflet des informations contenues dans les entretiens que des connaissances et théories mobilisées pour les organiser et les présenter. Ce travail de synthèse et de compte rendu explicatif s'appuie sur des citations extraites du corpus à l'appui des interprétations proposées. Il correspond à une tradition de recherche qualitative pure (Dumez, 2013) et peut s'envisager dans la perspective de la théorie

ancrée (Glaser et Strauss, 1967 ; Walsh, 2017). Dans leur ouvrage fondateur, Miles et Huberman (2003) élargissent le point de vue et décrivent la recherche qualitative comme une condensation progressive organisant la confrontation entre les idées du chercheur (théories, modèles) et les informations apportées par le terrain (entretiens, archives...).

2.1.2. Analyse de contenu

Le courant de l'analyse de contenu (Berelson, 1952 ; Drisko et Maschi, 2016) propose une approche plus formalisée et systématique de cette prise de connaissance. Elle repose sur l'identification de passages du corpus que le chercheur distingue en fonction de leur signification dénommée par des codes. Selon la granularité du découpage en unités de significations et le niveau de détail des codes, ce travail peut s'effectuer au niveau global de l'entretien considéré comme un tout exprimant la synthèse de ce que dit l'interviewé, ou au niveau plus fin des tours de parole ou des phrases. De nombreux travaux décrivent cette approche selon le niveau de détail du code et son intention (Saldaña, 2013) ou selon la manière de faire évoluer la codification (Gioia et al., 2012).

Le travail d'interprétation du chercheur se trouve ainsi encadré et systématisé (Krippendorff, 2013) et sa présentation objectivée (Bardin, 1977). Mais demeure la subjectivité irréductible de la lecture ou de la codification et la lourdeur du processus quand le corpus est très volumineux.

2.1.3. Analyse de contenu assistée par ordinateur

Longtemps effectué à la main, le travail de repérage et d'annotation peut être assisté par ordinateur avec des logiciels, appelés CAQDAS (*Computer Assisted Qualitative Data Analysis Software*), qui sont apparus dans les années 80. Les plus répandus répondent à l'approche de Miles et Huberman (2003). C'est le cas de NVivo et d'ATLAS.ti, dont le cœur consiste à repérer des extraits significatifs et à les affecter à des systèmes de codification évolutifs, aidant ainsi le chercheur à organiser ses idées et à rendre compte des thèmes qu'il découvre. D'autres logiciels comme QDminer, MaxQDA, Hyper Base, Sphinx, Alceste ou Iramuteq sont des outils statistiques d'exploration ou de *text mining*. Ils utilisent la statistique pour rendre compte des contenus, de leur organisation interne et de l'influence des contextes. Enfin, reprenant la tradition des thésaurus informatisés (Wordnet, Tropes), l'ingénierie linguistique a ouvert la voie à l'analyse sémantique et à la reconnaissance automatique des contenus, des opinions ou des sentiments (Goddard, 2011).

Malgré les progrès rapides des logiciels d'ADT, leur usage reste assez limité et le recours à Nvivo ou Atlas TI beaucoup plus fréquemment cité ne s'accompagne pas d'une contribution toujours visible sur les résultats obtenus dans l'analyse des corpus. Bref, tout se passe comme si les méthodes de recherche qualitative n'étaient que marginalement affectées par l'évolution des techniques logicielles et par les contributions de l'ADT.

2.2. L'analyse de données textuelles (ADT)

A l'origine développée pour l'analyse de grands corpus littéraires ou historiques, l'analyse de données textuelles naît de l'idée de traiter un texte comme des données constituées de mots et structurées selon les règles de la syntaxe et l'organisation du discours (parties, paragraphes et phrases). Si la répétition des mots fait sens (Lebart et al., 2020 ; Moscarola, 2018a), elle donne une première indication sur les contenus, précisée par la manière dont ils sont associés les uns aux autres. Ainsi la seule considération des fréquences d'utilisation des termes du corpus et de leur association peut-elle donner une idée du contenu des réponses aux entretiens.

Enfin, la considération des éléments de contexte et les différences qu'elle peut mettre en évidence dans l'usage des mots conduit à qualifier l'influence des locuteurs et des contextes. Ainsi, en traitant statistiquement les entretiens comme des ensembles de mots, on peut mettre en évidence des mots clés, des configurations thématiques ou des spécificités contextuelles.

2.2.1. ADT et approche exploratoire

Au cours des trente dernières années, les logiciels ont évolué grâce aux progrès des techniques de classification (Fenelon, 1999), de l'ingénierie linguistique (Veronis, 2004), et de l'analyse sémantique (Goddard, 2011). Les algorithmes de classification (Reinert, 1983) ont beaucoup contribué à la popularité des logiciels de l'ADT¹. Leur puissance s'est enrichie des apports de la lemmatisation, des thésaurus et de l'analyse sémantique (Baulac et al., 2014). Ils permettent d'identifier non seulement les mots clés, les expressions et les catégories thématiques révélées par leurs associations, mais également les concepts auxquels ils renvoient. Sur ces bases, la description des contenus peut être complétée par l'analyse de leur articulation interne et par leur dépendance par rapport aux influences du contexte.

Ces résultats, obtenus de manière automatique et reproductible, ne nécessitent aucune lecture ou codification préalable. Ils sont le reflet statistique des structures lexicales et sémantiques, présentées sous forme de nuages de mots, de cartes de configurations lexicales ou sémantiques ou de graphes d'influence. Leur exposition peut résumer en quelques pages l'intégralité du corpus. A l'analyste de savoir les lire pour leur donner sens en les interprétant. Mais ces « substituts lexicaux ou sémantiques » (Moscarola, 2018a) ne sont qu'une approximation et peuvent être trompeurs en conduisant le chercheur à projeter sur ce qu'il voit des interprétations erronées. Le retour au texte est donc indispensable pour vérifier par la lecture des entretiens ce que les substituts semblent indiquer.

2.2.2. Data visualisation et infographie dynamique

Les logiciels de dernière génération offrent une expérience de lecture hypertextuelle stimulée par l'examen des substituts lexicaux produits par l'ADT. La meilleure manière d'en saisir la portée pour l'analyse des entretiens est de l'expérimenter sur les données du cas qui suit.

Ce [lien](#) pointe sur une page contenant un nuage de mot clé, une catégorisation thématique, un nuage de concepts. En regard de ces substituts lexicaux et sémantiques figure le verbatim des entretiens découpés en phrases. En cliquant sur un des éléments des substituts, on lit le verbatim correspondant. Cela permet de vérifier le sens attribué aux mots clés, catégories thématiques ou concepts sélectionnés. Cet accès au corpus en donne une vision différente, à la fois très synthétique et détaillée, et constitue une autre manière de découvrir le contenu des entretiens.

2.2.3. ADT ciblée

L'exploration peut être poursuivie en ciblant plus précisément la recherche pour caractériser statistiquement le déroulement des entretiens, l'attitude des répondants, et l'influence de leur identité sur ce qu'ils disent.

Caractériser les entretiens et mettre en évidence les influences du contexte

Le nombre d'échanges, la longueur des questions et des réponses donnent une indication sur le déroulé des entretiens. De nombreux échanges, des réponses courtes indiquent que le répondant ne s'exprime que sous la pression de l'interviewer, alors que les interlocuteurs motivés s'expriment longuement de manière plus autonome. Plus les questions sont longues et

¹ Initialement implantés dans Alceste puis Iramuteq et Sphinx

nombreuses, plus grande est l'influence de l'interviewer. La variation de ces indicateurs peut être utilisée pour sélectionner les meilleurs entretiens.

D'autres indicateurs lexicaux permettent de qualifier l'énonciation et l'implication par l'usage des pronoms personnels notamment (e.g. Gavard-Perret et Moscarola, 1996).

Enfin, la possible influence de l'identité des interviewés, de leur statut et des contextes dans lesquels ils évoluent est observable par les variations dans l'usage des mots, des expressions et des catégories thématiques. Elles définissent des zones de langages (Moscarola 2018), et permettent de se prononcer sur l'influence des statuts ou des contextes. Et tout ceci sans lecture préalable !

Lire, coder et modéliser

Cependant l'exploration bute vite sur deux types de limites. D'une part, elle peut ne renvoyer que des évidences et ne rien apprendre de nouveau au chercheur qui connaît bien ses entretiens. Elle lui permet certes de vérifier ses impressions, de les établir objectivement pour ses lecteurs, mais sans aller plus loin. D'autre part, tant de pistes peuvent s'ouvrir au gré des idées suggérés par la data visualisation et les multiples croisements que le chercheur peut vite s'y perdre. Il doit être capable pour progresser d'organiser ses découvertes et sa pensée.

Les méthodologies traditionnelles de l'analyse de contenu apportent une réponse : lire, coder et organiser ; et l'ADT peut contribuer à ce travail. Si le chercheur n'a pas de modèle a priori (Boughzala et al., 2016), son travail consistera dans un premier temps à sélectionner les extraits qu'il regroupera par similitude comme on le fait en coloriant des passages dans un texte. Il peut les recopier dans les colonnes d'un tableur, les affecter aux nœuds de NVivo ou aux thèmes d'ATLAS.ti et de DataViv. Dans un deuxième temps, il cherche à donner un nom aux couleurs, colonnes, nœuds ou thèmes et, si nécessaire, à les regrouper pour parvenir à des catégories plus synthétiques. Il élabore ainsi, à partir des passages sélectionnés dans le texte, la liste des codes qui en rendent compte. L'ADT peut contribuer à ce travail en offrant la possibilité de lire le corpus par entrées lexicales ou sémantiques. Le chercheur se concentre ainsi sur les passages qu'il considère, à partir des substituts lexicaux, comme pertinents. Le résultat de ce travail est d'établir la liste des thèmes/codes correspondants aux extraits sélectionnés et, lorsque la lecture a été exhaustive, d'en faire le décompte pour indication de leur importance.

Pour aller plus loin et comprendre ce que disent les répondants, le chercheur peut regrouper les codes en catégories/dimensions ou thèmes plus généraux (Gioia et al., 2012) et distinguer des causes et des effets. Il construit ainsi un modèle articulant les thèmes évoqués dans les entretiens. Ce modèle peut être testé par régression linéaire ou équations structurelles sur la base de variables évaluant pour chaque entretien l'évocation plus ou moins forte des éléments du modèle. On construit pour cela des intensités thématiques définies comme la proportion dans un entretien des extraits correspondant au thème considéré par rapport au nombre total d'extraits, tous thèmes confondus. On fait ainsi l'hypothèse de travail que, plus un thème est évoqué (relativement aux autres thèmes), plus il compte pour celui qui parle. Un exemple de cette démarche est présenté dans le cas qui suit. Il met en évidence une autre contribution de l'analyse de données textuelle utilisée pour quantifier des données qualitatives après lecture et codification.

La construction de dictionnaires thématiques

On peut s'affranchir de la lecture des entretiens en construisant un thésaurus². Cela revient à dresser des listes de termes de significations voisines ou évoquant le même thème. On procède à partir du lexique avec la possibilité de vérifier le sens des mots dans leur contexte. Chaque liste reçoit le nom du concept qu'elle évoque. Ces listes sont utilisées pour calculer l'intensité du champ lexical des concepts correspondants. Cela revient à coder automatiquement les entretiens. C'est l'ordinateur qui lit et qui attribue selon la granularité choisie le poids relatif de chaque thème.

2.2.4. Choisir une approche

Exploration ou analyse de contenu ciblée, le choix d'une approche dépend de l'orientation par une question de recherche plus ou moins précise. Dans tous les cas, la description des propriétés statistiques et lexicales du corpus, qui est contrôlée par la datavisualisation et le retour au texte, contribuent à un exposé objectif et critique du matériau analysé. Mais le recours à l'ADT ne se justifie vraiment que pour des corpus d'une taille suffisante pour que les effets de répétition soient statistiquement significatifs. Dans la tradition des études quantitatives, trente entretiens sont nécessaires et, selon la théorie des actes de langage (Austin, 1962), c'est le nombre de mots du corpus qu'il faut considérer³.

Le chercheur doit-il s'engager lui-même dans un travail de lecture et de codification? Cela dépend de la taille du corpus et des catégories conceptuelles auxquelles le modèle de recherche renvoie. La codification peut prendre beaucoup de temps. Si les codes renvoient à des faits circonscrits et repérables par des listes de mots, on peut construire un thésaurus. S'ils impliquent une interprétation ou une appréciation, on ne peut faire autrement que de passer à la lecture. Quitte à mener l'analyse sur un échantillon du corpus comme cela a été fait dans le cas qui suit où un peu plus de la moitié des entretiens ont été codés.

Enfin, notons une possibilité intéressante lorsque le chercheur transcrit lui-même les entretiens qu'il a effectués. A la fin du travail de la transcription, il documente une fiche dans laquelle il indique dans quelle mesure cet entretien répond aux différents aspects du modèle, comme s'il répondait aux questions de recherche à la place de l'interviewé. En procédant ainsi, le chercheur engage sa subjectivité⁴, qu'il justifie en sélectionnant des verbatims illustratifs. Mais il peut revendiquer la position d'expert que lui confère sa connaissance du terrain et des théories auxquelles il se réfère. Dans le cas étudié ci-dessous, ce travail a été fait a posteriori à des fins de triangulation.

3. Illustration : un cas de changement organisationnel

Examinons l'application de ces méthodes sur le corpus des 40 entretiens d'une recherche en gestion faite grâce au logiciel Sphinx (Moscarola, 2018b) et à Dataviv (visualisation de données). Les liens permettent au lecteur de faire l'expérience de ce corpus et des méthodes utilisées ([Lien vers la première application utilisée par les chercheurs](#)).

3.1. L'étude d'un changement organisationnel

3.1.1. Le cas Alpha

2 Un thésaurus est une arborescence de concepts définis par des ontologies (liste de mots) et des réseaux sémantiques (Moscarola 2014)

3 Le logiciel Sphinx met en garde l'utilisateur quand le corpus fait moins de 5000 mots pleins.

4 Beaucoup plus qu'en codant des passages particuliers, car il est amené à porter une appréciation globale.

Alpha est une entreprise de plus de cent ans. Ce cabinet de services professionnels spécialisé en propriété intellectuelle opère en Australie et ses avocats représentent des clients partout dans le monde. Alpha est régi par un accord de partenariat signé par 33 associés, principalement des avocats en brevets.

Afin de répondre aux défis de l'industrie et de compenser la perte de revenus traditionnels, Alpha aspire à offrir une gamme de services plus large à ses clients. A cette fin elle a recruté pour créer de nouveaux services hébergés au sein du partenariat et a engagé une équipe marketing pour créer une image différenciante. Cela impliquait un exercice de repositionnement de marque majeur. La nouvelle identité « déclarée » d'Alpha le positionne comme un partenaire commercial holistique fournissant des services globaux en matière de propriété intellectuelle.

Cependant, après trois ans, les services ajoutés n'ont pas été à la hauteur de leur promesse et pratiquement aucun effet de levier sur la liste des clients d'Alpha n'a été atteint. Notre objectif est donc, de manière exploratoire, d'observer ce changement et, de manière ciblée, de comprendre les éléments qui concourent à la réalisation d'une nouvelle aspiration identitaire.

3.1.2. Les entretiens et le corpus

Les entretiens en face-à-face en langue anglaise ont duré entre 0,5 h et 1,5 h et ont été menés à la manière d'une conversation visant à recueillir la perception de l'expérience vécue par les personnels par rapport à l'évolution de la nature et des exigences de l'industrie de la propriété intellectuelle, et comment elle affecte les pratiques de travail. Les 40 membres de l'entreprise interrogés reflètent la variété des fonctions au sein du groupe. Les entretiens ont tous été intégralement retranscrits et validés par les répondants. Chaque entretien comporte une moyenne de 50 échanges questions-réponses. Les réponses font 25 mots en moyenne pour un corpus total de 52 472 mots.

3.1.3. Une analyse en deux temps

Dans un premier temps, il a été procédé à une exploration lexicale et sémantique des 40 entretiens afin de faire émerger les principaux thèmes, catégories significatives et concepts. Dans un second temps, les entretiens ont été lus et codés en vue de pouvoir répondre aux objectifs de la recherche : comprendre la formation des aspirations identitaires. Face à la lourdeur de la tâche, l'analyse a été limitée aux entretiens restituant le mieux la diversité des personnes interrogées du point de vue de leur position et de l'apport à la question de recherche (22 entretiens sur 40).

3.2. Exploration lexicale et sémantique

3.2.1. Aperçu lexical et sémantique

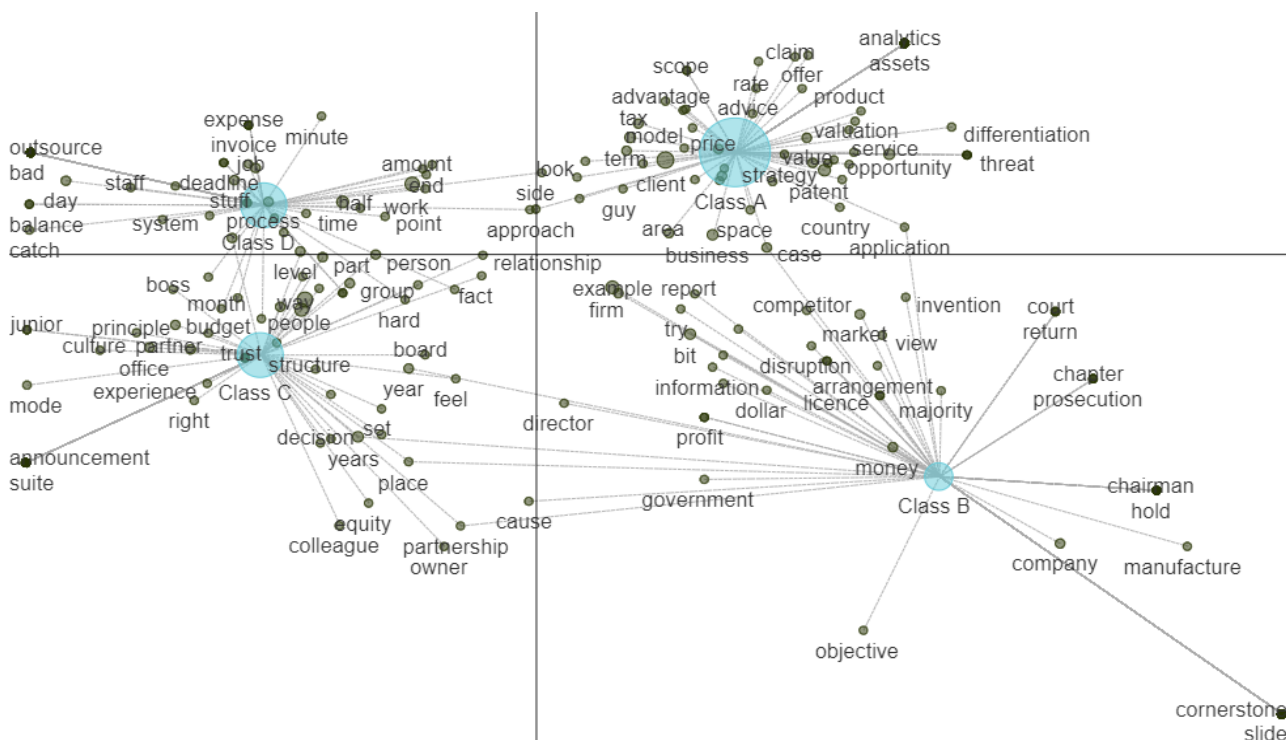
Les cinq mots clés les plus fréquents (*client, people, work, good, firm*) révèlent les préoccupations des répondants. Cela est précisé par la typologie du corpus découpé en phrases (8826) qui permet de distinguer trois groupes de préoccupations : les gens au travail, les clients et la pression du temps, et la firme sur son marché. L'application des catégories d'un thésaurus standard, malgré son éloignement du sujet considéré pointe sur l'importance de l'engagement économique (*Wage earners*) et sur une certaine inquiétude (*Fright and hope*). Ces résultats d'une première approximation lexicale et sémantique sont accessibles avec le lien [Mots clés](#). Il permet un premier niveau d'interprétation.

3.2.2. Interprétation dans la perspective de la recherche

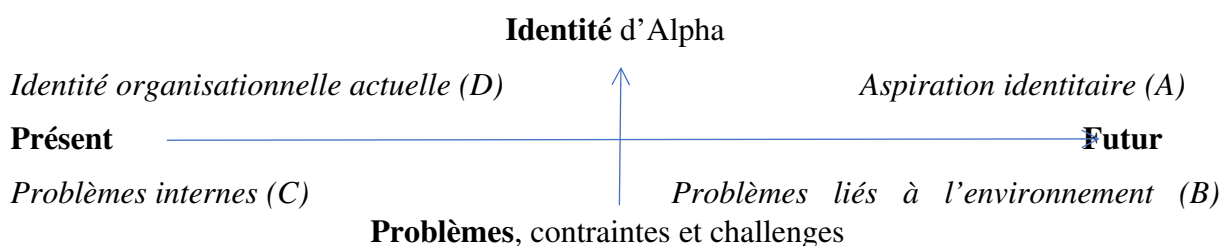
La figure 1 ci-dessous, visible dans le lien [Classification et AFC](#), montre les quatre groupes émergeant d'une séquence d'analyses factorielles des correspondances multiples. Elle met en évidence les mots sur représentés dans chaque classe.

Le premier groupe (classe A en haut à droite) est centré autour des mots « *service* », « *advice* », « *attorney* », « *client* ». Le deuxième groupe (classe B) en bas à droite met en évidence les termes « *market* », « *competitor* », « *disruption* », « *money* » et « *profit* ». Le troisième groupe en bas à gauche (classe C) est centré autour du mot « *structure* », avec également « *right* », « *decision* », « *partner* », « *people* », « *board* ». Enfin le quatrième groupe en haut à gauche (classe D) regroupe les termes « *time* », « *work* », « *job* », « *staff* », « *deadline* ».

Figure 1 : Représentation AFC de la Classification



A la lueur des termes de chaque classe et de la recherche exploratoire en termes de changement organisationnel et d'aspiration identitaire, l'interprétation de ces clusters peut se faire de la manière suivante : l'axe des abscisses peut être intitulé « Présent - Futur », celui des ordonnées « Contraintes par rapport à l'identité - Identité ».



En poussant l'interprétation un cran plus loin, cette classification fait nettement ressortir que les personnes interrogées s'expriment sur des thèmes différents et peuvent être distinguées de manière plus globale en deux catégories. D'un côté, les personnes ancrées dans le présent qui

parlent de l'identité actuelle d'Alpha en mettant l'accent sur les fondements de cette identité : activité d'écriture de brevets très contraignante en termes de coûts et prenant beaucoup de temps, les avocats étant très sensibles à la qualité de ce qu'ils produisent (même si le coût facturé au client, en heures de travail, ne pourra couvrir le temps passé). Ces acteurs évoquent donc aussi naturellement les contraintes qui pèsent sur leur travail, liés essentiellement à la structure de partenariat qui est le propre de l'entreprise (et souvent des entreprises de services professionnels) : les partenaires sont au conseil d'administration et ce sont eux qui prennent les décisions stratégiques. Certains reconnaissent toutefois que cette structure est très inhibante et contraignante dans la mesure où les partenaires ont du mal à renoncer à leur liberté et à leur manière de faire traditionnelle, sans contrôle. Aussi toute mesure visant à introduire des processus plus formalisés n'est pas bien acceptée au sein d'Alpha, car elle augmente le contrôle de leurs activités.

De l'autre côté, on trouve au contraire les personnes qui s'expriment plutôt sur ce qu'elles souhaiteraient faire à l'avenir, et la manière dont elles espèrent voir Alpha évoluer, notamment en termes de développement d'activité et d'organisation. Ces aspirations vers une gamme de services étendues allant du conseil en taxation de R&D au conseil en matière de propriété intellectuelle vont de pair avec les changements de l'environnement (marchandisation, digitalisation, concurrence accrue, etc.) qui représentent un danger pour le métier de base d'Alpha et pèsent sur ses marges et profits.

3.2.3. L'influence des contextes

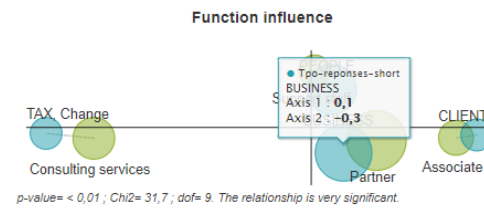
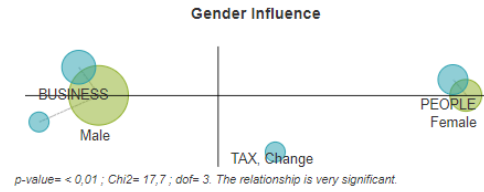
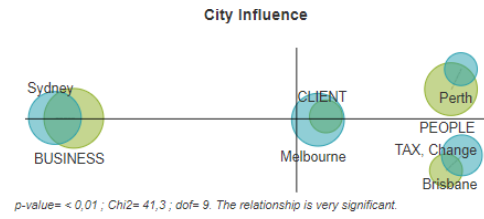
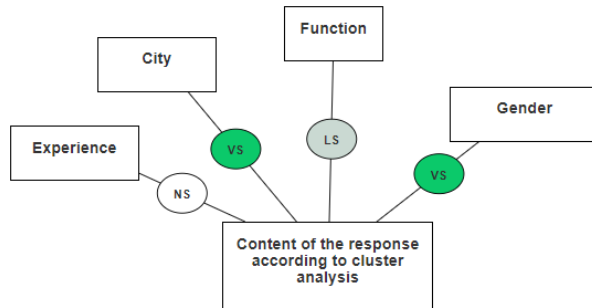
Le lien [Contexte](#) met en évidence l'influence de l'identité du répondant sur la thématique de ses réponses établie par une classification automatique. Il montre l'influence de certaines variables de contexte sur le contenu des réponses. Selon le test du Chi², la ville d'exercice et le genre jouent de manière très significative ($p < 0.01$). A Sydney, le thème du *business* (service aux clients, valeur à apporter, stratégie et marchés) est surreprésenté, alors qu'à Melbourne les acteurs interrogés s'expriment plus en termes d'orientation client par rapport à leur métier de base d'avocat. En ce qui concerne le genre, on voit que les femmes mettent l'accent sur les employés et l'importance accordée aux personnes alors que les hommes parlent plus du contenu de leur activité. Enfin, en termes de fonction, dont la relation avec le contenu présente un degré de significativité moindre, on voit que les personnes appartenant à la branche Conseil (Consulting services) insistent sur le thème du changement.

La figure 2 ci-dessous résume ces influences :

Figure 2 : Facteurs de contingence

According to the chi2 test, the city, the group of practice and the gender have an influence on the contents of the answers :

In Sydney the theme of business is over represented, women are distinguished by their focus on the theme of people and the theme of change is characteristic of consulting services



3.3. Analyse de contenu ciblée

Afin de préciser et d’approfondir la problématique du changement et de l’identité mise en évidence par l’approche exploratoire, une lecture systématique du corpus a été entreprise. La lourdeur de la tâche a conduit à se limiter à 22 entretiens, comme indiqué dans la partie méthodologique.

3.3.1. Lire et coder

Les entretiens ont tout d’abord été codés de manière émergente et inductive, en sélectionnant les extraits renvoyant explicitement à la question de recherche. Ceci a conduit à la sélection de 510 extraits repérés par une codification très fine (210 codes). Ces codes ont tout d’abord été regroupés afin d’obtenir, à la manière de la méthode « à la Gioia » des catégories de premier ordre, soit 20 au total. Ces 20 catégories ont été regroupées, de manière thématique par rapport au questionnement théorique sur les freins à la réalisation d’une aspiration identitaire.

3.3.2. Construire un modèle

Quatre thèmes ou dimensions de second ordre ont été définis : Identité organisationnelle, Problèmes internes représentant autant de contraintes et challenges à surmonter, Problèmes externes liés à la complexité institutionnelle, et Aspiration identitaire. Ils constituent le modèle théorique proposé pour expliquer la formation de l’aspiration identitaire.

Tableau 1. Thèmes du modèle et codification de premier ordre

Aspiration	Identité organisationnelle actuelle	Problèmes externes	Problèmes internes
Holistic-IP	Profession	Cost pricing	Markets
Better processes	Partnership	Resistance of client	Commoditization
Reputation	Individual	Management	Competition
Resources	Relationships	Unstructured processes	Patent changes
Value pricing	Atmosphere»	Cultural	

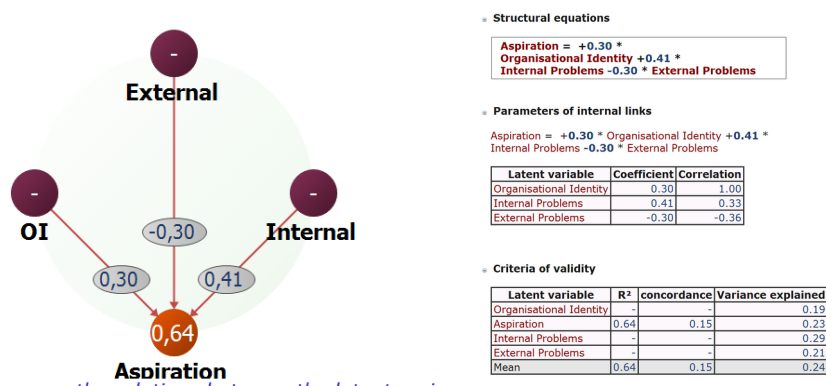
Les codes de premier ordre (cf. tableau 1) associés de manière univoque à chacun des 503 extraits caractérisant le contenu des 22 entretiens ont été utilisés pour calculer l'intensité avec laquelle ils sont évoqués dans chacun des entretiens (Nombre d'évocations d'un code / nombre total des évocations). On procède ainsi à une évaluation du contenu des entretiens sous la forme de 20 variables.

3.3.3 Tester par un modèle d'équations structurelles

Ces variables sont utilisées comme variables manifestes dans un modèle d'équation structurelle dont les variables latentes sont les thèmes de second ordre définis ci-dessus. Le test porte sur l'explication des évocations de l'aspiration considérée, de manière formative, comme résultant des évocations de l'identité organisationnelle, et des problèmes internes et externes. Le modèle obtenu comme résultat (cf. figure 3) est excellent dans la mesure où il restitue 64% de la variance.

L'interprétation de ce modèle peut être la suivante : les dimensions ne sont pas intrinsèques et indépendantes. En effet, on observe une très forte contingence entre les différentes dimensions et une relation explicative forte de l'évocation des problèmes internes et externes et de l'identité organisationnelle. Les signes positifs indiquent que lorsque les personnes parlent d'aspiration, elles mentionnent aussi les aspects identitaires et internes. En revanche, lorsque les personnes interrogées parlent de logiques externes, elles ne rêvent pas et ne s'expriment pas en termes d'aspiration (identitaire, mais aussi en termes organisationnel, d'activité, etc.). Comme si les répondants percevaient les bénéfices/implications/contraintes internes de l'aspiration identitaire mais n'arrivaient pas à articuler les bénéfices et contraintes pour et sur son environnement. En d'autres termes, il est difficile de se projeter vers une aspiration et une logique que l'on ne comprend pas.

Figure 3 : Relations entre les variables du modèle (équations structurelles PLS)



3.3.4. Interpréter

3.3.5. Contrôle par une approche globale

La démarche précédente est très analytique : repérage des extraits significatifs, codification détaillée de premier ordre. Elle présente l'avantage d'une codification au plus près du texte : les extraits codés sont courts et l'interprétation facilement contrôlable. On calcule ensuite pour chaque entretien la fréquence relative d'évocation des codes considérée comme représentative des contenus.

Une manière de contrôler cela est de confronter l'approche détaillée à une codification globale des entretiens, directement effectuée au niveau des variables de second ordre ([Grille](#)

codification). Cette évaluation engage beaucoup plus fortement la subjectivité du chercheur qui se trouve conduit à exprimer une impression générale. Mais il peut l'enrichir de sa connaissance du terrain et de toutes les autres informations qui ne figurent pas dans le texte et qui l'aident à formuler ce jugement.

Le test par régression linéaire du modèle ainsi documenté donne un bon résultat ($r=0,8$) et, comme pour les résultats obtenus par la codification détaillée, le poids des problèmes internes est prépondérant. Mais le sens des autres influences se trouve inversé (cf. [Lien](#)).

4. Discussion

Nous avons montré avec cet exemple que les corpus d'entretiens peuvent être abordés de manière exploratoire par les techniques de l'analyse de données textuelle, lexicale et sémantique. Cette première approche du corpus pose l'analyse statistique comme préalable à la lecture et à l'interprétation des substituts lexicaux qu'elle produit. Le retour au texte et la datavisualisation permettent de contrôler cette lecture qui, *in fine*, donne un sens aux structures révélées par l'ADT.

Mais l'usage de la quantification peut également intervenir en second lieu après lecture et codification du corpus. L'analyse des données textuelles consiste alors à mettre en œuvre des techniques pour quantifier les contenus en référence à des modèles⁵ dont les variables sont analysées selon les méthodes quantitatives classiques. Dans ce cas, la lecture précède la quantification.

Ainsi, quelle que soit l'approche (inductive ou déductive), les entretiens peuvent être analysés par les techniques de l'analyse de données textuelles. Mais ce dépassement des clivages traditionnels ne fait pas disparaître l'essence de la recherche qualitative : construire le sens par la lecture attentive et critique des textes et par celle, plus technique, de substituts lexicaux ou des variables quantitatives déduites de la codification.

La mise en œuvre des logiciels que nous avons utilisés, offre une expérience de la prise de connaissance des corpus renouvelée par la visualisation de données. Partagée avec le lecteur elle lui donne aussi, dans l'esprit critique du débat scientifique, la possibilité de discuter les interprétations proposées par l'auteur.

Demeure l'obstacle des habitudes. A commencer par les normes de la communication écrites encore peu ouverte aux extensions web. S'y ajoutent les habitudes du clivage quali/quant et la technicité des notions et méthodes de l'ADT, qui requiert des compétences spécifiques. Enfin, ces techniques, par leur efficacité de premier abord et par l'aisance qu'elle donne à ceux qui les maîtrisent, comportent aussi le risque réel d'usages abusifs ou trompeurs si la construction du sens se fait en dépit du temps, long, de la lecture et de la culture.

5. Bibliographie

Austin, J.L. 1962. *How to Do Things with Words*. Oxford University Press, Oxford.

Bardin, L. 1977. *L'analyse de contenu*. PUF, Paris.

Boughzala Y, Moscarola J., Hervé M. 2016. *Sphinx Quali : un nouvel outil d'analyses textuelles et sémantiques*. Journées d'analyse statistique des données textuelles (JADT), Nice.

⁵ Le système des codes et de leurs relations.

- Benzecri, J.P. 1975. L'analyse des données. *Population*, 30^e année, n°6, 975-1017.
- Berelson, B. 1952. *Content analysis in communication research*. The Free Press, New York.
- Drisko, J.W., Maschi, T. 2016. *Content analysis*. Oxford University Press, Oxford.
- Dumez, H. 2013. *Méthodologie de la recherche qualitative*. Vuibert, Paris.
- Fénelon, J.P. 1999. *Qu'est-ce que l'analyse de données ?* Lefonen, Paris.
- Gavard-Perret, M-L., Moscarola, J. 1996. *Lexical analysis in marketing: discovering the contents of the message or recognizing the models of enunciation*. French-German Workshop on New Developments and Approaches in Consumer Behavior Research (1996), University of Potsdam, Germany, 40 - 59.
- Gioia, D.A., Corley, K.G., Hamilton, A.L. (2012). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, vol. 16, n°1, 15-31.
- Glaser, B.G., Strauss, Anselm L. 1967. *The discovery of grounded theory: strategies for qualitative research*. Chicago: Aldine.
- Goddard, C. 2011. *Semantic Analysis: A Practical Introduction*. Oxford University Press, Oxford.
- Krippendorff, K. 2013. *Content Analysis. An Introduction to Its Methodology* (3rd edition). Sage Publications, California.
- Lebart, L., Pincemin, D., Poudart, C. 2020. *Analyse des données textuelles*. Presses de l'Université du Québec, Canada.
- Miles, M. B., Huberman, M. A. 2003. *Analyse des données qualitatives* (2^{ème} édition). De Boeck, Paris.
- Moscarola, J. 2018a. *Faire parler les données*. Editions EMS, France.
- Moscarola, J. 2018b. Visualisation de données et infographie dynamique : le logiciel Sphinx. In F. Chevalier, M. Cloutier, N. Mitev (eds), *Méthodes de recherche pour le DBA*. EMS, Paris (p. 71-94).
- Moscarola, B., Moscarola, J. 2018. *ADT et visualisation, pour une nouvelle lecture des corpus Les débats de 2^{ème} tour des Présidentielles (1974-2017)*. Journées d'analyse statistique des données textuelles (JADT), Rome.
- Paillé, P. 2006. *La méthodologie qualitative - Postures de recherche et travail de terrain*. Armand Colin, Paris.
- Reinert, A. 1983. Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, tome 8, n°2, 187-198.
- Saldaña, J., 2013. *The coding manual for qualitative researchers* (2^{ème} édition). Sage, London.
- Véronis, J. 2004. Le traitement automatique des corpus oraux. In *Traitement automatique des langues*, vol. 45, n°2. Hermès, Lavoisier, Paris.
- Walsh, I. 2017. *Discovering new theories*. Editions EMS, France.