



HAL
open science

La quantification des données qualitatives : intérêts et difficultés en sciences de gestion

Isabelle Royer, Lionel Garreau, Thomas Roulet

► **To cite this version:**

Isabelle Royer, Lionel Garreau, Thomas Roulet. La quantification des données qualitatives : intérêts et difficultés en sciences de gestion. Finance Contrôle Stratégie, 2019, 2019 (NS6), 10.4000/fcs.3312 . hal-02303982

HAL Id: hal-02303982

<https://hal.science/hal-02303982>

Submitted on 2 Oct 2019

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

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

La quantification des données qualitatives : intérêts et difficultés en sciences de gestion

Isabelle Royer, Lionel Garreau et Thomas Roulet



Édition électronique

URL : <http://journals.openedition.org/fcs/3312>

DOI : 10.4000/fcs.3312

ISSN : 2261-5512

Éditeur

Association FCS

Ce document vous est offert par Bibliothèque de l'Université Paris-Dauphine



Référence électronique

Isabelle Royer, Lionel Garreau et Thomas Roulet, « La quantification des données qualitatives : intérêts et difficultés en sciences de gestion », *Finance Contrôle Stratégie* [En ligne], NS-6 | 2019, mis en ligne le 23 mai 2019, consulté le 17 septembre 2019. URL : <http://journals.openedition.org/fcs/3312> ; DOI : 10.4000/fcs.3312

Ce document a été généré automatiquement le 17 septembre 2019.

Tous droits réservés

La quantification des données qualitatives : intérêts et difficultés en sciences de gestion

Isabelle Royer, Lionel Garreau et Thomas Roulet

Introduction

- 1 Depuis plusieurs années, on peut noter le développement de techniques mobilisant la quantification des données qualitatives, telles que l'analyse qualitative comparée (QCA), l'analyse des réseaux sociaux, l'analyse textuelle, ainsi que l'utilisation accrue de nouvelles données qualitatives quantifiées issues d'internet. Il est difficile d'affirmer que la proportion de la quantification en recherche au cours des dernières décennies augmente, car cela demanderait une analyse systématique des publications. Pourtant, des observations vont dans ce sens, tant en économie et sciences sociales au cours du 20ème siècle (Chiapello et Desrosières, 2006), que dans le domaine de la recherche en gestion où l'on note une augmentation de la formalisation mathématique et de la quantification (Maurand-Valet, 2011). Ces développements sont suffisamment importants pour qu'il soit intéressant de dresser un bilan de l'intérêt et des questions relatives à la quantification.
- 2 La quantification consiste à associer une valeur numérique à un objet d'étude (Espeland et Stevens, 2008). Elle comprend deux aspects principaux : le comptage et la mesure. Alors que le comptage concerne des objets ou événements similaires, la mesure consiste à assigner un nombre à certains aspects des objets ou événements selon une règle ou convention (Stevens, 1968) et pas aux objets eux-mêmes. Comme l'indiquent Pedhazur et Pedhazur Schmelkin (1991), on peut compter des individus mais pas les mesurer en tant qu'entité, on ne peut mesurer que leur taille, leur anxiété, leur habileté, etc. De la même manière, on pourra compter des entreprises et mesurer leur chiffre d'affaires, leur retour sur investissement, leur nombre de salariés par exemple. Si le comptage est important en stratégie et en finance, les chercheurs en gestion sont principalement intéressés par la mesure. Étant donné la quantité d'attributs d'un individu ou d'une organisation, seule la

théorie permet de déterminer ce qu'il est pertinent de quantifier relativement à son objet d'étude (Pedhazur et Pedhazur Schmelkin, 1991).

- 3 Toutes les données qualitatives, qu'elles soient sous forme de texte ou d'image, peuvent être quantifiées et utilisées comme variables d'un modèle testé avec des techniques statistiques dès lors que leurs conditions d'utilisation sont satisfaites. L'exemple le plus simple est le traitement des questions ouvertes d'un questionnaire qui sont codées en variables binaires (qui prennent les valeurs 1 et 0 pour présence et absence respectivement) puis insérées dans les analyses avec les variables quantitatives du même questionnaire. Au-delà de cet exemple simple et ancien, la quantification est utilisée aujourd'hui dans plusieurs démarches autrefois typiquement qualitatives telle que l'étude de cas et met en œuvre des technologies de plus en plus sophistiquées incluant le *machine learning* où le chercheur apprend au programme à coder comme il le souhaite jusqu'à l'obtention de résultats satisfaisants pour produire des variables utilisables dans des approches quantitatives.
- 4 La quantification, qu'elle consiste à compter ou mesurer, est une démarche qui peut demander une grande quantité de travail et de ressources (Espeland et Stevens, 2008). Elle est utilisée dès lors que le chercheur considère que la quantification présente un intérêt pour son travail. Ses avantages techniques sont importants mais son utilisation est inutile dans certaines situations (Bardin, 2013).
- 5 En plus d'une démarche technique, la quantification – en tant que production et communication de nombres – a des implications significatives sur la science, l'organisation et plus largement la société si bien que certains souhaitent en faire un nouveau domaine de la sociologie (Diaz-Bone et Didier, 2016). Se basant sur l'approche du langage d'Austin, Espeland et Stevens (2008) considèrent les nombres comme des mots et les statistiques comme faisant partie de la grammaire qui aide à leur donner du sens. « *Comme des actes de locution, les nombres impliquent des conventions de production, expression et interprétation qui les rendent intelligibles ou pas* » (Espeland et Stevens, 2008 : 403). Cette dimension sociologique de la quantification peut procurer d'autres avantages et être source de questionnement.
- 6 Cet article vise à introduire la démarche de quantification des données qualitatives en stratégie, management et finance. Une première partie présente les intérêts qui peuvent exister à quantifier, que ceux-ci soient de nature instrumentale, symbolique ou esthétique. La seconde se focalise sur les questionnements et choix qui ponctuent la démarche de quantification depuis le choix du matériau à quantifier, de l'énumération, du recours à l'automatisation jusqu'aux objectifs de la recherche et aux considérations éthiques. Les articles du numéro spécial sont ensuite introduits.

Les intérêts de la quantification

- 7 La quantification présente plusieurs avantages que l'on peut classer en fonction de leur nature. Elle présente des intérêts méthodologiques, augmente la crédibilité et constitue actuellement une opportunité.

Intérêts méthodologiques

- 8 Le principal avantage de la quantification est la commensuration (Espeland et Stevens, 1998). En effet, une mesure commune crée des relations spécifiques entre les objets, laquelle permet de les représenter plus facilement et surtout de les comparer relativement à l'aspect mesuré (Espeland et Stevens, 1998). Par exemple, 20 kilos et 40 kilos sont plus précis et plus facile à comprendre et comparer qu'une expression verbale qui pourrait y être associée telle que lourd et très lourd (Pedhazur et Pedhazur Schmelkin, 1991). La comparaison est plus ou moins précise en fonction de la mesure utilisée. On distingue habituellement quatre types de mesures qui sont de la plus simple à la plus élaborée : nominale, ordinale, intervalle et ratio (Pedhazur et Pedhazur Schmelkin, 1991). Ces mesures sont hiérarchisées au sens où les propriétés de la dernière (le ratio) incluent celles de la précédente (l'intervalle), et ce jusqu'à la mesure élémentaire (la mesure nominale), base de toutes les autres reflétant la catégorie d'appartenance.
- 9 Un second avantage de la quantification réside dans la facilité d'analyse des nombres comparés aux données qualitatives, ce qui ouvre une large variété d'analyses réalisables rapidement. La quantification permet de construire des proxys pour des caractéristiques difficiles à capter. Surtout, le nombre permet de condenser et réduire l'information de telle manière qu'elle puisse être traitée mécaniquement par un logiciel de statistiques (Espeland et Stevens, 1998). La quantification permet ainsi « *des transferts, des comparaisons, des agrégations, des manipulations standardisées par le calcul, et des interprétations routinisées* » (Desrosières, 2012 : 269). Par exemple Bernard *et al.* (2018) y ont eu recours pour étudier la capacité des éditos des rapports de développement durable à prédire la performance RSE. Les traitements possibles à mettre en œuvre dépendent du type de mesure utilisé. Bien que les mesures nominales offrent moins de possibilités que les mesures plus élaborées, elles permettent déjà un grand nombre de traitements. Le temps de traitement d'un logiciel variant peu en fonction de la quantité de données, plus le nombre d'observations est important, plus il est intéressant de quantifier les données qualitatives. Plus on souhaite multiplier les analyses, plus il est intéressant de quantifier. En effet, l'analyse d'un jeu de données quantifiées étant rapide comparé à celui des données non quantifiées, il est particulièrement efficient de quantifier lorsqu'on désire utiliser plusieurs fois les mêmes données dans des analyses différentes.
- 10 Un troisième avantage est la transparence des analyses qui sont explicites et peuvent donc être vérifiées et répliquées (Langley, 1999). Cette transparence confère une fiabilité aux analyses valorisée dans de nombreuses perspectives de recherche. Cette transparence constitue aussi une incitation à une plus grande rigueur des analyses.

Crédibilité et communication

- 11 En plus de ces trois avantages de nature instrumentale, le nombre bénéficie d'une valeur symbolique qui n'est pas négligeable. Plusieurs auteurs en sociologie (e.g., Desrosière, 2012 ; Espeland et Stevens, 2008) et en management (e.g., Denis, Langley et Rouleau, 2006 ; Déjean, Gond et Leca, 2004) indiquent que les nombres ont une forte capacité de persuasion via la légitimité et le pouvoir. Dans le domaine académique, plusieurs sources de crédibilité peuvent être associées au nombre suivant celles répertoriées par Espeland et Stevens (2008). La première est leur validité en tant que représentation de ce qui est

mesuré indépendamment du chercheur, caractéristique d'une perspective réaliste. L'objectif est un parfait isomorphisme entre le nombre et l'aspect mesuré. Cette approche correspond à la théorie de la mesure qui inclut les imperfections (erreurs de mesure). La deuxième source de crédibilité vient de la capacité des nombres à offrir un langage commun, compréhensible, impersonnel et permettant donc de communiquer malgré des distances sociale, géographique ou disciplinaire (Espeland et Stevens, 2008). En absorbant l'incertitude, le nombre facilite la communication (March et Simon, 1958). Qualifié d'« *objectivité mécanique* » par Porter (1995), le nombre est particulièrement utile en situation conflictuelle car il est basé sur une règle impersonnelle. La troisième source de crédibilité est liée à leur usage. Plus les nombres sont utilisés et encastrés dans un réseau d'acteurs qui les produisent et les utilisent, plus ils sont crédibles. Cette crédibilité dans l'usage peut amener les chercheurs à mobiliser des exemples d'utilisation (de base de données par exemple) plutôt que justifier la pertinence des variables de la base de données utilisée pour leurs travaux. La quatrième source de persuasion est une image traditionnelle de rationalité et d'objectivité, associée traditionnellement aux recherches quantitatives.

- 12 Dans leur sociologie de la quantification, Espeland et Stevens (2008) consacrent une part importante à la dimension esthétique des nombres. Bien que cette dimension esthétique varie sensiblement en fonction du temps et des disciplines, Espeland et Stevens (2008) considèrent que deux critères sont quasiment universels : la clarté et la parcimonie qui invitent à n'indiquer que les éléments nécessaires et suffisants à la communication de l'information principale. Ces deux critères sont très présents en sciences de gestion dans les recommandations aux auteurs aussi bien sur le plan de la théorie que de la communication des résultats. Du fait de sa capacité à réduire et agréger l'information, le nombre offre une possibilité de clarté et de parcimonie sans égale comparée à des données qualitatives. Il suffit de comparer les tableaux d'analyses des articles empiriques quantitatifs et qualitatifs pour s'en convaincre. Le développement par Fiss (2011) d'une présentation des résultats de la méthode QCA sous forme de cercle n'est pas étranger à cette esthétique qui rend la lecture des résultats plus aisée.

Opportunités

- 13 Le développement des *big data* constitue une opportunité sans précédent du recours à la quantification des données qualitatives. De grandes quantités de données sous forme de texte, d'image ou de vidéo sont accessibles relativement aisément sur les sites Web, les blogs, Twitter, Instagram pour ne citer que quelques exemples. De par leur facilité d'accès et leur très grand nombre, ces données se prêtent particulièrement à la quantification. Les recherches en management et finance ont commencé à intégrer ces données textuelles (par exemple : Wang *et al.*, 2016 ; Renault, 2017). Les *big data* comportent également d'autres types de données, notamment visuelles qui demeurent encore peu exploitées en science de gestion. Les analyses de contenu visuel sont encore peu fréquentes. On peut toutefois citer par exemple des travaux sur le genre à partir de l'analyse de rapports annuels (Duff, 2011). Bell et Davidson (2013) voient dans les données visuelles un nouveau tournant des recherches sur les organisations qui ne peut que se développer. L'accessibilité à des données qualitatives va de pair avec le développement de programmes et de logiciels permettant de les exploiter de plus en plus amplement. Les logiciels spécialisés en analyse de contenu tels que Sphinx Lexica, Spad-T, Alceste ou

Tropes et en lexicométrie tels que Prospero se développent ainsi que des modules sur plateforme libre, notamment R (voir Maire et Liarte dans ce numéro pour une comparaison des logiciels). L'analyse automatisée des images par *machine learning* dans d'autres disciplines ouvre la voie vers l'intégration de données nouvelles dans nos disciplines via la quantification.

Questionnements et choix liés à la quantification

- 14 Le passage de données qualitatives à des variables qui pourront faire l'objet de traitements statistiques implique certains points de vigilance et est sujet à des choix qui sont présentés ici.

Le matériau à quantifier

- 15 La quantification de données qualitatives étant destinée à des analyses quantitatives ou a minima automatisées, la constitution de l'échantillon, ou encore du corpus pour les analyses textuelles, nécessite de suivre les critères requis pour les traitements choisis. Par suite, la taille de l'échantillon requis et sa structure peuvent varier très fortement. Chaque méthode d'analyse présente des critères qui lui sont propres mais on retrouve des points communs : l'homogénéité, la représentativité, et l'exhaustivité. Ces trois critères sont nécessaires pour donner du sens aux résultats quantitatifs. L'échantillon doit être suffisamment homogène (techniques de collecte identiques, individus comparables) afin que les comparaisons entre individus fassent sens. Ces individus doivent représenter un ensemble, ou une population selon le vocabulaire des études quantitatives. Lorsque la taille de la population est petite, celle-ci pourra être étudiée en intégralité. Lorsqu'elle est grande, un échantillon aléatoire ou représentatif peut être utilisé (voir Royer et Zarlowski, 2014). La représentativité peut conduire à ce que l'hétérogénéité soit un facteur clef dans la sélection des sources afin de représenter la diversité d'opinions existantes. C'est en particulier le cas pour la constitution d'un corpus de média (Roulet, 2015 ; Roulet et Clemente, 2018). Enfin, l'information pour chaque individu doit être exhaustive afin de ne pas affecter les traitements.
- 16 Du fait des critères d'homogénéité, de représentativité et d'exhaustivité, certaines méthodes de recherche qualitatives ne se prêtent pas à la quantification. C'est par exemple le cas de la théorie enracinée dans sa version originale (Glaser et Strauss, 1967) où il est possible d'étudier des cas de façon partielle dans l'objectif de saturer les catégories. La différence de complétude en termes de recueil de données conduit à un manque d'homogénéité entre les cas qui ne permet pas de les comparer quantitativement. De manière plus générale, les études qualitatives exploratoires ou intensives (Grawitz, 2000) se prêtent peu à la quantification puisqu'elles n'ont vocation à être ni représentatives, ni exhaustives.
- 17 Par ailleurs, les méthodes d'analyse par quantification des occurrences de thèmes (Bardin, 2013) dans des entretiens doivent être maniées avec précaution. En effet, le comptage d'occurrences de thèmes peut se révéler très influencé par la manière de mener les entretiens (Mbengue *et al.*, 2014). Dans ce cadre, on peut souligner qu'il existe actuellement dans les recherches utilisant le comptage d'occurrences une certaine opacité sur la manière dont les entretiens sont menés et ses potentielles conséquences sur les résultats obtenus.

- 18 Enfin, la quantification s'adresse à des échantillons ou corpus de taille moyenne ou grande. Par exemple, la mise en œuvre de l'algèbre booléenne de la méthode QCA n'est pas utile si toutes les informations peuvent être résumées sur une page. Par suite, la technique a longtemps été recommandée pour des échantillons de taille intermédiaire de 11 à 50 (Chanson *et al.*, 2005) avant que d'autres auteurs en montrent la performance sur des échantillons plus grands (Misangyi *et al.*, 2017). De même, lorsque le nombre de cas est limité (Langley, 1999) ou le corpus réduit (Bardin, 2013), il est préférable de garder toute la richesse des cas dans leur contexte pour améliorer l'analyse plutôt que d'utiliser la quantification, qui est par nature réductrice.

Produire le nombre

- 19 La quantification de données qualitatives requiert non pas une mais deux opérations. La première consiste en un codage et la seconde en une énumération (Bardin, 2013). Le codage consiste à faire des inférences à partir de quelque chose, souvent du texte. Pour cela, une première question consiste dans le choix de l'unité de codage (ou unité d'enregistrement). Il y en a principalement trois : le mot (ou groupe de mots), la phrase et le sens appelé aussi thème (Allard-Poesi, 2003 ; Weber, 1990). Pour ce dernier, la taille est variable puisqu'il s'agit d'extraire une portion de texte suffisante pour qu'elle ait une signification non ambiguë. De ce fait, quelques mots parfois suffisent alors que dans d'autres cas un long paragraphe peut être nécessaire. Ce choix a un impact sur le comptage : plus l'unité est grande, plus les fréquences sont faibles.
- 20 Chaque unité de codage est associée à une catégorie par inférence. De nouvelles questions se posent ici concernant la définition des catégories. En particulier : les catégories sont-elles mutuellement exclusives ou non ? (Weber, 1990). Roulet (2019) crée par exemple une série de catégories pour dépeindre les différentes formes d'hostilité à l'égard des pratiques des banques d'investissement dans les médias écrits, et ce sur la base d'un codage qualitatif d'un échantillon d'articles. Des catégories non exclusives réduisent l'éventail des techniques quantitatives possible à utiliser pour les analyses. Une seconde question est le caractère étroit ou, au contraire, large de la catégorie (Weber, 1990). Là aussi l'incidence est directe. Une catégorie définie de manière large aura des fréquences plus élevées que si elle est définie de façon plus étroite.
- 21 L'énumération est l'association d'une valeur numérique à la catégorie choisie. Plusieurs règles d'énumération existent dont certaines diffèrent selon la méthode d'analyse choisie. Pour l'analyse de contenu par exemple, Bardin (2013) indique la présence, la fréquence, la fréquence pondérée, l'intensité, la direction, l'ordre, la co-occurrence. L'utilisation de ratios (nombre de mots sur quantité de texte) peut permettre de contrôler par la taille du texte. D'autres ratios sont possibles, par exemple l'étendue de la couverture média pour une organisation (Roulet, 2019). Un même codage produira des valeurs différentes selon la règle d'énumération choisie. Par exemple, si sur un ensemble de 50 entretiens on trouve 100 itérations du code A et 50 itérations du code B, il sera différent de l'analyser sur le fait que : a) le code A est présent 100 fois ; b) le code A apparaît 2 fois plus que le code B ; c) le code A apparaît au moins une fois dans 48 des 50 entretiens ; d) le code A apparaît 51 fois dans 1 entretien et 1 fois uniquement dans les 49 autres entretiens ou encore e) dans 60 % des cas, le code B apparaît en cooccurrence avec le code A. Toutes ces analyses amènent à des interprétations possibles différentes qu'il convient de clarifier tant dans l'exposé de la méthode que dans celui des résultats. Ainsi,

ce n'est pas parce que les logiciels permettent la quantification relativement simple et un ensemble d'analyses que le chercheur doit fonctionner comme le logiciel le permet : il doit bien garder à l'esprit son projet et ne pas se fourvoyer du fait des options informatiques disponibles (Bandeira-de-Mello et Garreau, 2011). Par ailleurs, si la présence est le codage habituel (qu'on peut exprimer 1/0) on trouve aussi parfois un codage des niveaux d'intensité (faible, moyen et fort par exemple) en particulier pour les études de cas (par exemple Martin et Eisenhardt, 2010). L'analyse de réseaux sociaux pour sa part est riche d'indicateurs spécifiques (Lazega, 1994).

Quantification manuelle ou automatique ?

- 22 La distinction entre codification manuelle et automatique traite de l'usage de programmes informatiques dans le codage des données jusqu'à l'attribution d'une valeur. Actuellement aucun logiciel ne permet de réaliser seul une quantification pertinente sans préparation préalable des données et paramétrage des traitements. Nous entendons ici par quantification manuelle un codage qui repose uniquement sur le chercheur qui infère un sens aux données. Ce codage manuel s'effectue aujourd'hui à l'aide de logiciels non spécialisés ou spécialisés tels que Atlas-ti ou N-Vivo¹ qui permettent de faire du comptage mais sont insuffisants pour la quantification automatisée. Le codage automatisé est ici défini comme le résultat du programme spécifié par le chercheur puis réalisé par la machine sous contrôle du chercheur.
- 23 La quantification automatique présente des avantages. Elle augmente la rapidité de l'analyse (Bardin, 2013). Elle force la rigueur et la fiabilité puisque tout doit être explicité et spécifié dans le programme (Bardin, 2013 ; Weber, 1990). Par suite, la reproduction dans une perspective cumulative ou collaborative est facilitée (Bardin, 2013 ; Weber, 1990). Ces avantages énoncés pour les analyses de contenu peuvent être généralisés assez largement à d'autres méthodes.
- 24 Dans le cadre des analyses de contenu, Bardin (2013) recommande d'utiliser l'ordinateur lorsque l'unité d'analyse est le mot et l'indicateur fréquentiel. Pour des unités de codage de grande taille, supérieure au paragraphe, notamment un texte entier, lorsque l'analyse est complexe et comporte un grand nombre de variables à traiter simultanément, lorsqu'on souhaite effectuer une analyse des co-occurrences, lorsque la recherche implique plusieurs analyses successives et lorsque des opérations statistiques complexes sont prévues, l'automatisation, à l'inverse, est déconseillée.
- 25 Une difficulté des logiciels pour l'analyse textuelle réside dans le traitement des mots et des sigles ambigus qui seront regroupés alors que leurs significations peuvent être différentes, des locutions inconnues qui seront traitées comme des mots isolés et des phrases complexes que l'ordinateur ne classe pas toujours correctement. Par ailleurs, chaque logiciel présente des particularités liées à son concepteur qui favorisent certaines analyses au détriment d'autres (voir par exemple Maire et Liarte, ainsi que Blanc, Peton et Garcias dans ce numéro). En conséquence, le choix du logiciel adapté à son approche requiert de l'attention.
- 26 En cas de codage automatisé de textes se pose aussi la question de l'utilisation d'un dictionnaire général préexistant (Weber, 1990). Ces dictionnaires généraux comportent des catégories préétablies associées à des mots ou locutions de plusieurs mots et expressions idiomatiques (Bardin, 2013). L'intérêt des dictionnaires est multiple : ils fournissent une palette de catégories parmi lesquelles choisir ; ils réduisent le temps de

construction, validation et révision du dictionnaire final ; ils standardisent la classification, permettent la comparaison entre études (Weber, 1990). Weber (1990) cite notamment Harvard IV Dictionary et Lasswell Value Dictionary (LVD) pour la langue anglaise. Malheureusement, ces dictionnaires généraux ne sont pas nécessairement adaptés au langage technique dont les termes sont parfois absents mais surtout dont la signification peut être différente. Ainsi, Wang, Wezel et Forgues (2016) qui ont travaillé sur des réponses d'hôteliers à leurs clients ont pu utiliser les dictionnaires du logiciel LIWC qui donnent des résultats similaires à un codage manuel. Même si les dictionnaires généraux Harvard IV et Loughan and McDonald Dictionary offrent une première approche quantitative qui peut compléter une approche qualitative (Roulet, 2015), Renault (2017) montre qu'ils ne sont pas adaptés à l'analyse des tweets des investisseurs en raison d'un style spécifique à la contrainte de 140 caractères et de la signification particulière de certains mots et groupes de mots dans le domaine financier. Par suite, l'auteur a constitué un dictionnaire spécifique en utilisant le *machine learning*. Dans la même lignée, Picault et Renault (2017), ont développé un dictionnaire spécifique pour appréhender les significations des mots utilisés par la Banque Centrale Européenne afin que les analyses reflètent le contexte spécifique de ce type de communication. Développer des dictionnaires spécifiques peut se faire via des approches qualitatives et inductives (Roulet, 2019). Nous entrevoyons donc ici une tendance où des méthodes spécifiques sont développées pour appréhender les contextes d'utilisation des mots pour une analyse ensuite automatisée.

Quantifier pour faire quoi ?

- 27 La quantification n'est pas un objectif à part entière mais un moyen de réduire l'information pour faciliter des analyses automatiques. Dès lors que les données prennent la forme de variables, qu'elles soient nominales, ordinales ou d'intervalles, elles ouvrent l'univers très vaste des techniques statistiques. Parmi cet univers des possibles, on peut toutefois distinguer deux catégories d'objectifs : la mise en évidence de régularités et le test de variables. La mise en évidence de régularités ou patterns est une tradition des recherches qualitatives (Langley, 1999) et inductives. La comparaison de cas (avec leurs points communs et différences) vise principalement à identifier les similarités en termes de structure ou de processus. La méthode QCA de Ragin (1987) permet par l'algèbre booléenne d'identifier les configurations permettant d'aboutir à un résultat. Par exemple Lehiany et Chiambaretto (dans ce numéro) analysent 11 alliances ferroviaires et identifient 4 configurations aboutissant à la stabilité d'une alliance. La méthode des événements d'Abbott (1990) permet d'identifier des phases d'un processus et le gamma de Pelz (1985) permet d'ordonner les étapes d'un processus. Les approches inductives d'analyse des réseaux sociaux visent à identifier la structure relationnelle (Lazega, 1994). Ces approches inductives mettent souvent en œuvre une analyse globale des cas étudiés, et les éléments récurrents ou structurels mis en évidence par l'analyse peuvent constituer une contribution à eux seuls.
- 28 Le second objectif est le test de variables. Dans cette deuxième perspective hypothético déductive, le recours à des données qualitatives peut se limiter à un moyen de collecte de données jugé plus pertinent pour opérationnaliser des concepts à tester, encore mal définis a priori. Les questions ouvertes des enquêtes entrent dans cette catégorie. Aujourd'hui, les *big data* ouvrent des possibilités de constitution de bases de données

d'opinions mais aussi de comportements autrefois difficilement accessibles. Par exemple, Renault (2017) a étudié l'impact des sentiments des investisseurs sur le réseau spécialisé *StockTwist* sur les rendements du marché boursier. Pour cela, il a utilisé 5 années de messages, soit plus de 59 millions de messages codés automatiquement en sentiments positifs ou négatifs après une procédure de *machine learning*. Wang, Wezel et Forgues (2016) se sont intéressés à la manière dont les organisations répondent à une sous-évaluation par leurs clients. Pour cela, ils ont recueilli toutes les réponses des hôteliers sur TripAdvisor de la région de Londres sur 10 années, soit 209 000 évaluations pour 596 hôtels, et les ont traitées à l'aide du logiciel LIWC. Utilisant l'analyse de réseaux de manière déductive, Cattani et Ferriani (2008) ont étudié le rôle des réseaux sociaux sur la créativité individuelle. Pour cela, ils se sont intéressés aux principaux membres des équipes des films produits par les 8 majors américaines durant 12 ans figurant dans la base de données IMDB de l'industrie cinématographique. Les deux types d'objectifs peuvent se combiner, les patterns ayant émergé de l'approche inductive étant ensuite testés.

La responsabilité du chercheur dans la quantification

- 29 Plusieurs sociologues mentionnent la rétroaction de la quantification (Desrosière, 2012 ; Espeland et Stevens, 2008). La quantification conduit à penser et agir différemment. C'est notamment le cas des indicateurs de performance qui conduisent les agents à modifier leur comportement (Déjean, Gond, et Leca, 2004). Ces conséquences ne sont toutefois pas toujours conçues par le chercheur, ni même imaginées. Certaines quantifications ont des conséquences qui sont différentes voire opposées à celles souhaitées par le chercheur. Espeland et Stevens (2008) citent notamment l'exemple de Kinsey qui a défini l'homosexualité en tant que comportement et pas élément constitutif d'une personne et a abouti à un mouvement politique de catégorisation de personnes. Les mesures parfois initialement prévues pour décrire peuvent également ensuite être utilisées pour d'autres finalités telles que l'évaluation ou le contrôle (Espeland et Stevens, 2008). Les sources du matériau quantifié peuvent aussi s'avérer biaisées : par exemple quantifier les sources médias implique une prise en compte des approches politiques que ces derniers peuvent avoir et il faut donc équilibrer les sources afin de présenter une quantification objective d'une donnée qualitative hétérogène quant à son orientation (Roulet et Clemente, 2018). Du fait de son fort impact potentiel sur le monde social, la quantification invite le chercheur à se poser des questions relatives à sa responsabilité sociale. Sans être un frein à sa recherche, il est possible de réfléchir aux utilisations possibles de son travail et informer des conséquences (Durand *et al.*, 2009 ; Royer, 2011).

Introduction des articles

- 30 Ce numéro spécial réunit quatre manuscrits couvrant différents aspects de la quantification des méthodes qualitatives. Tout d'abord, deux manuscrits portent sur le raffinement et l'exploitation de la méthode d'analyse qualitative comparée (QCA) visant à comparer des cas sur la base de causes identifiées ou conditions. Chanson (2019) offre une contribution méthodologique concernant le choix des catégories retenues pour l'analyse QCA, améliorant la fiabilité de cette méthode. Chanson défend et enrichit l'idée d'une « condensation » des données, en se fondant sur les potentialités du « *crispy set* » QCA. Il

propose l'utilisation d'hypothèses simplificatrices pour identifier des catégories pertinentes. Lehiany et Chiambaretto (2019) mobilisent pour leur part la méthode QCA pour examiner empiriquement la stabilité des alliances. Leur approche méthodologique leur permet d'identifier différentes combinaisons causales, et les auteurs montrent comment la méthode QCA permet de d'appréhender la complexité causale et ainsi de dépasser les limites des modèles linéaires.

- 31 Les deux autres manuscrits présents dans ce numéro portent sur l'analyse textuelle et la profondeur que la quantification des données qualitatives peut apporter dans ce contexte. Maire et Liarte (2019) nous proposent un cadre méthodologique pour faire émerger des catégories thématiques automatiquement à partir d'une base textuelle, nous rappelant, dans un autre contexte, l'idée de condensation proposée par Chanson (2019). Maire et Liarte (2019) offrent une approche en trois temps – classification, représentation et labellisation – et passent en revue les outils qui permettent cette démarche. Leur approche est ensuite appliquée, à titre d'illustration, aux calendriers des Scouts et Guide de France de 1936 à 2016. Enfin, le travail de recherche de Blanc, Peton et Garcias (2019) se propose d'examiner les macro-discours culturels, c'est-à-dire les éléments discursifs à la fois constitutifs et supports des croyances et pratiques des acteurs sociaux. Dès lors que l'intérêt du concept est établi, la question de la capture empirique de cet objet discursif apparaît cruciale. Les auteurs détaillent une approche lexicométrique sans préconception, permettant l'émergence quasi-inductive des thématiques clefs de ces macro-discours. En utilisant le logiciel Prospéro, ils illustrent leur propos en observant les macro-discours autour du nucléaire.

Conclusion

- 32 Comme dans toute science, la qualité et la pertinence des connaissances apportées par les sciences de gestion sont intimement liées à des enjeux de méthode. Les méthodes qualitatives ont vu leur niveau d'exigence augmenter au cours des dernières décennies, tout en restant fidèles à leur philosophie d'origine. La quantification de données qualitatives, qui est dès l'origine de l'analyse de contenu un des objectifs affirmés (Bardin, 2013), se voit aujourd'hui prendre de nouvelles dimensions de par les avancées technologiques tant dans les méthodes de recueil que de traitement des données. Ainsi des phénomènes qui étaient par nature appréhendés de façon qualitative comme le langage (Graebner *et al.*, 2012) sont aujourd'hui adressés de façon de plus en plus courante avec des traitements quantitatifs ou avec des traitements mixtes.
- 33 Pourtant, de nombreuses voies de recherche peuvent encore être explorées pour améliorer la qualité et la pertinence des méthodes de quantification. Les limites et difficultés de la quantification présentées dans cet article constituent une partie des voies de recherches possibles. De plus, chaque technique ou catégorie de techniques (analyse textuelle, analyse de réseau, QCA, ...) offre des questions spécifiques qui requièrent probablement des réponses elles aussi spécifiques. D'autres limites sont pour leur part inhérentes à la quantification, dont la réduction des données conduit à un appauvrissement (Graebner *et al.*, 2012). Cette perte de richesse des données est considérée comme pertinente lorsqu'elle peut être compensée par des avantages de nature différente, qu'ils soient de nature instrumentale ou symbolique.

- 34 Le contexte des *big data* constitue toutefois à la fois une opportunité sans précédent de quantification de données qualitatives du fait de leur accessibilité mais aussi un challenge du fait de leur volume qui requiert l'automatisation du codage. Si des logiciels existent aujourd'hui pour traiter de façon satisfaisante des corpus de textes homogènes, le traitement automatisé de matériaux hétérogènes et non textuels relève de l'exploration. Les erreurs de classement du fait de l'ambiguïté de mots, sigles, locutions ou constructions grammaticales complexes sont accrues lorsque les matériaux textuels sont hétérogènes. Par conséquent, le développement de logiciels permettant une meilleure prise en compte du contexte ou de techniques permettant d'identifier des passages potentiellement problématiques nécessitant une intervention humaine offre des voies de recherche prometteuses. La quantification de données visuelles reste quant à elle à explorer. Plusieurs axes de recherches en plein développement tels que le tournant affectif (Clough et Halley, 2007), et désormais le tournant matériel invitent à l'utilisation de données visuelles, si bien que certains annoncent un tournant visuel (Boxenbaum *et al.*, 2018). La quantification de ces données visuelles, notamment automatisées, ouvre elle aussi de vastes perspectives en sciences de gestion.

BIBLIOGRAPHIE

- Abbott, A. (1990). A Primer on Sequence Methods. *Organization Science*, vol. 1, n° 4, p. 375-392.
- Allard-Poesi, F. (2003). « Coder les données » in Y. Giordano (Ed.), *Conduire un projet de recherche : une perspective qualitative*. EMS.
- Bandeira-de-Mello, R., & Garreau, L. (2011). « L'utilisation d'Atlas.ti pour améliorer les recherches dans le cadre de la Méthode de la Théorisation Enracinée (MTE) : panacée ou mirage ? », *Recherches Qualitatives*, vol. 30, n° 2, p. 175-202.
- Bardin, L. (2013). *L'analyse de contenu*, PUF.
- Bell, E., & Davison, J. (2013). « Visual management studies : Empirical and theoretical approaches », *International Journal of Management Reviews*, vol. 15 n° 2, p. 167-184.
- Bernard Y., Godard L., Hervé F. & Zouaoui M. (2018), « Les éditos des rapports de développement durable servent-ils à quelque chose ? Une étude empirique de leur capacité à prédire la performance RSE », *Finance Contrôle Stratégie*, vol. NS-4
- Blanc A., Peton H., & Garcias F. (2019). « L'analyse lexicométrique des macro-discours par les vocabulaires – enjeux théoriques et méthodologiques », *Finance Contrôle Stratégie*.
- Boxenbaum, E., Jones, C., Meyer, R. E., & Svejenova, S. (2018). « Towards an Articulation of the Material and Visual Turn in Organization Studies », *Organization Studies*, vol. 39, n° 5-6, p. 597-616.
- Cattani, G., & Ferriani, S. (2008). « A Core/Periphery Perspective on Individual Creative Performance : Social Networks and Cinematic Achievements in the Hollywood Film Industry », *Organization Science*, vol. 19, n° 6, p. 824-844.
- Chanson, G. (2019). « Pour une comparaison systématique de cas : l'apport de csQCA à la condensation des données qualitatives », *Finance Contrôle Stratégie*

- Chanson, G., Demil, B., Lecocq, X., & Sprimont, P.-A. (2005). « La place de l'analyse qualitative comparée en sciences de gestion », *Finance Contrôle Stratégie*, vol. 8, n° 3, p. 29-50.
- Chiapello, E., & Desrosières, A. (2006). « La quantification de l'économie et la recherche en sciences sociales : paradoxes, contradictions et omissions. Le cas exemplaire de la positive accounting theory » in Eymard-Duvernet, F. (Ed.), *L'économie des conventions, méthodes et résultats*, La Découverte, p. 297-310.
- Clough, P., & Halley, J. (Eds.). (2007). *The Affective Turn : Theorizing the Social*, Duke University Press.
- Déjean, F., Gond, J.-P., & Leca, B. (2004). « Measuring the unmeasured : An institutional entrepreneur strategy in an emerging industry ». *Human Relations*, vol. 57 n° 6, p. 741-764.
- Denis, J.-L., Langley, A., & Rouleau, L. (2006). « The power of number in strategizing », *Strategic Organization*, vol. 4, n° 4, p. 349-377.
- Desrosières, A. (2012). « Est-il bon, est-il méchant ? Le rôle du nombre dans le gouvernement de la cité néolibérale », *Nouvelles perspectives en sciences sociales : Revue internationale de systémique complexe et d'études relationnelles*, vol. 7, n° 2, p. 261-295.
- Diaz-Bone, R., & Didier, E. (2016). « Introduction : The Sociology of Quantification-Perspectives on an Emerging Field in the Social Sciences », *Historical Social Research/Historische Sozialforschung*, p. 7-26
- Duff, A. (2011). « Big four accounting firms' annual reviews : A photo analysis of gender and race portrayals », *Critical Perspectives on Accounting*, vol. 22, n° 1, p. 20-38.
- Durand, R., Charreire-Petit, S., & Warnier, V. (2009). *Tribune. Revue Française de Gestion*, vol. 4, p. 15-28.
- Espeland, W. N., & Stevens, M. L. (1998). « Commensuration as a Social Process », *Annual review of sociology*, vol. 24, n° 1, p. 313-343.
- Espeland, W. N., & Stevens, M. L. (2008). « A sociology of quantification », *European Journal of Sociology/Archives Européennes de Sociologie*, vol. 49 n° 3, p. 401-436.
- Fiss, P. C. (2011). « Building better causal theories : A fuzzy set approach to typologies in organization research », *Academy of Management Journal*, vol. 54 n° 2, p. 393-420.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory : Strategies for Qualitative Research*, Aldine De Gruyter.
- Graebner, M. E., Martin, J. A., & Roundy, P. T. (2012). « Qualitative data : Cooking without a recipe », *Strategic Organization*, vol. 10, n° 3, p. 276-284.
- Grawitz, M. (2000). *Méthodes des sciences sociales*, Dalloz.
- Langley, A. (1999). « Strategies for Theorizing from Process Data », *Academy of Management Review*, vol. 24, n° 4, p. 691-710.
- Lazega, E. (1994). « Analyse de réseaux et sociologie des organisations », *Revue française de sociologie*, p. 293-320.
- Lehiany, B. & Chiambaretto, P. (2019), « L'apport de la méthode fsQCA à l'étude de la complexité causale : une application à la stabilité des alliances », *Finance Contrôle Stratégie*.
- Maire, S. & Liarte, S. (2019). « Classifier, représenter et labelliser : Cadre et outils méthodologiques associés pour une émergence automatique de thématiques sur données textuelles », *Finance Contrôle Stratégie*.

- March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.
- Martin, J. A., & Eisenhardt, K. M. (2010). « Rewiring : Cross-business-unit collaborations in multi-business organizations », *Academy of Management Journal*, vol. 53, n° 2, p. 265–301.
- Maurand-Valet, A. (2011). « Choix méthodologiques en Sciences de Gestion : pourquoi tant de chiffres ? », *Management & Avenir*, vol. 3, p. 289-302.
- Mbengue, A., Vandangeon-Derumez, I., & Garreau, L. (2014). « Construire un modèle », in R. A. Thiétart (Ed.), *Méthodes de recherches en Management*, Dunod, 4^{ème} édition, p. 334–387.
- Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. (2017). « Embracing Causal Complexity : The Emergence of a Neo-configurational Perspective », *Journal of Management*, vol. 43, n° 1, p. 255-282.
- Pedhazur, E. J., & Pedhazur Schmelkin, L. (1991). *Measurement, design, and analysis : An integrated approach*. Hillsdale : Lawrence Erlbaum Associates.
- Pelz, D. C. (1985). « Innovation complexity and the sequence of innovating stages ». *Knowledge*, vol. 6 n° 3, p. 261-291.
- Picault, M., & Renault, T. (2017). « Words are not all Created Equal : A New Measure of ECB Communication », *Journal of International Money and Finance*, vol. 79, p. 136–156.
- Porter, T. M. (1992). « Quantification and the accounting ideal in science », *Social studies of science*, vol. 22, n° 4, p. 633-651.
- Ragin, C. C. (1987). *The comparative Method : Moving Beyond Qualitative and Quantitative Strategies*, University of California Press.
- Renault, T. (2017). « Intraday Online Investor Sentiment and Return Patterns in the US Stock Market », *Journal of Banking & Finance*, vol. 84, p. 25-40.
- Roulet, T. (2015). « “What good is Wall Street ?” Institutional Contradiction and the Diffusion of the Stigma over the Finance Industry », *Journal of Business Ethics*, vol. 130, n° 2, p. 389-402.
- Roulet, T. J., & Clemente, M. (2018). « Let’s Open the Media’s Black Box : The Media as a Set of Heterogeneous Actors and not only as a Homogenous Ensemble », *Academy of Management Review*, vol. 43, n° 2, p. 327-329.
- Roulet, T. J. (2019). « Sins for Some, Virtues for Others : Media Coverage of Investment Banks’ Misconduct and Adherence to Professional Norms during the Financial Crisis », *Human Relations*.
- Royer, I. (2011). « La responsabilité des chercheurs en gestion », *Revue Française de Gestion*, vol. 7, p., 65-73.
- Royer, I., & Zarlowski, P. (2014). « Echantillon(s) », in R.-A. Thiétart (Ed.), *Méthodes de recherche en management*, Dunod, 4^{ème} édition, p. 219-260) :.
- Stevens, S. (1968). « Measurement, statistics and the Schemapiric View », *Science*, vol. 161, n° 3844, p. 849-856.
- Wang, T., Wezel, F. C., & Forgues, B. (2016). « Protecting Market Identity : When and How Do Organizations Respond to Consumers’ Devaluations ? », *Academy of Management Journal*, vol. 59, n° 1, p. 135-162.
- Weber, R. P. (1990). *Basic content analysis*, vol. 49, Sage.

NOTES

1. Le site www.nerdyscholar.com présente de nombreuses fonctionnalités de l'utilisation du logiciel Nvivo, qu'elles soient associées à la quantification ou non.

RÉSUMÉS

Dans cet article introductif du numéro spécial sur la quantification des données qualitatives, nous présentons l'intérêt et l'ampleur du sujet pour le champ des sciences de gestion. Quantifier implique l'association de valeurs numériques – via la mesure ou le comptage – à des jeux de données pour lesquels ces valeurs ne sont pas évidentes – comme par exemple du texte ou des images. En premier lieu nous explorons les raisons qui justifient la quantification des données qualitatives. La transparence et la possibilité de comparer les jeux de données de manière formelle figurent parmi les principaux avantages de la quantification. La crédibilité et la facilité de communication en sont d'autres. Nous discutons ensuite des précautions à prendre. En particulier, il est important de sélectionner avec attention le matériau à quantifier, et de réfléchir à la production des valeurs numériques. Enfin nous notons la responsabilité du chercheur et le recul nécessaire pour une quantification rigoureuse.

AUTEURS

ISABELLE ROYER

Université de Lyon Jean Moulin, iaelyon business school, Magellan
isabelle.royer@univ-lyon3.fr

LIONEL GARREAU

Université Paris-Dauphine, PSL, DRM [UMR CNRS 7088], 75016 PARIS
lionel.garreau@dauphine.psl.eu

THOMAS ROULET

Cambridge University, Judge Business School et Girton College
t.roulet@jbs.cam.ac.uk