

Innovations dans un logiciel d'analyse qualitative de données : l'intégration d'outils de visualisation

Udo Kuckartz, Ph.D.

Philipps-Universität Marburg

Résumé

La présentation et la visualisation de données sont des domaines de la recherche qui deviennent de plus en plus importants, en particulier dans les approches qualitatives et mixtes. Cet article couvre ces thèmes à partir de deux perspectives. Premièrement, l'attention est portée sur les questions méthodologiques et théoriques afin de démontrer que la visualisation est créative et utile pour l'exploration et la découverte de logiques, de relations et de structures dans les données. Des scientifiques ont montré de façon concluante que les informations complexes et les corrélations peuvent être traduites en des représentations graphiques facilement lisibles. Deuxièmement, la partie principale de l'article s'intéresse aux pratiques de recherche, et en particulier à l'utilisation de nouveaux outils de visualisation intégrés dans des logiciels d'analyse de données qualitatives, tel que MAXQDA. Une distinction majeure est opérée entre deux types de visualisation : les visualisations internes à un cas et les visualisations multi-cas. La présentation visuelle des données a plusieurs fonctions dans le processus d'analyse : l'agrégation, la réduction et l'exploration des données, la gestion des idées ainsi que la présentation des résultats.

Mots clés

VISUALISATION, DONNÉES QUALITATIVES, LOGICIELS, MAXQDA

Introduction

Cet article prend une perspective qui est, dans une certaine mesure, la contrepartie du point de vue dominant de la sociologie du visuel. Mon attention ne se porte pas sur le visuel, ou l'image, en tant qu'objet de la recherche et de la méthodologie en sciences sociales. Le visuel est approché plutôt de l'autre côté : je vais de la recherche vers le visuel. Plus spécifiquement, je me demande comment visualiser les connexions rencontrées à travers l'analyse qualitative de données et le gain que cela peut apporter au processus d'analyse qualitative.

Le sujet de ce papier concerne la recherche en tant qu'activité productrice d'images, par exemple la transformation de résultats de recherche et des processus de recherche dans le monde des images, tels que les formes, graphes, modèles, carte conceptuelle, etc. Ce sujet est présent dans mon esprit depuis plusieurs années et j'aimerais présenter une série de visualisations récemment développées qui ont été entretemps mises en œuvre à travers le logiciel d'analyse qualitative MAXQDA.

La visualisation a gagné en importance dans de nombreuses disciplines scientifiques tant comme outil d'évaluation que comme mode de présentation des résultats d'analyse. Les sciences sociales sont, par contre, une branche des sciences dans laquelle le travail et les arguments sont basés principalement sur des textes.

Contrairement aux sciences naturelles ou à la médecine, où il est de pratique courante de présenter les résultats de façon visuelle et, de plus en plus pour les diagnostics médicaux, en utilisant des techniques d'imagerie, la méthodologie en sciences sociales utilise ces possibilités de façon plus que modérée. Il existe très peu de littérature en méthodologie des sciences sociales qui traitent explicitement de possibilités de modélisation des résultats d'analyse (Banks, 2001; Blasius & Greenacre, 1998). Au mieux, nous le rencontrons plus régulièrement dans le cadre d'analyse de réseaux (par exemple : Krempel, 2005). La raison, telle que Feldmann (2003) la présente, est une aversion générale des sciences sociales envers les images. La sociologie est un enfant de la rationalité et des lumières, et est très sceptique quant aux images de par son héritage.

The holy Sociology writings of Marx, Durkheim, Weber, Parsons, Luhmann, Habermas, Giddens etc. are mostly image-less, seldom defiled with statistics. You should not make any images! Yet, not only the classics, but also the newer sociological writings that use older and newer media as well as the internet are largely image-free (Feldmann, 2003).

Visualisation et perception humaine

Dans son film concernant le changement climatique, Al Gore exprime ainsi que « le non-expert peut « voir » les relations et est impressionné par sa modélisation ». En effet, les représentations graphiques nous permettent de présenter des faits complexes de façon à les rendre facilement et rapidement compréhensible par tous. Seuls quelques scientifiques en sciences sociales se sont intéressés à ce sujet et à la question de comment configurer au mieux ces représentations.

Lothar Krempel représente une exception dans sa dissertation à propos de la visualisation en réseau, en ce qu'il a également abordé l'histoire de la visualisation de l'information (Krempel, 2005) : les cartes géographiques, dit Krempel, existent déjà en 2000 A.J.-C. et il est de notoriété que la cartographie a joué un rôle central dans le transport maritime. Tout qui voudrait savoir comment se rendre d'un endroit à un autre peut trouver facilement en regardant une représentation graphique. Imaginons un instant que nous devons transcrire en mots toute l'information contenue dans la carte du métro londonien (*The Tube*). Ou encore si, en tant qu'utilisateur du métro, nous avons à lire un texte qui décrit les connexions et les stations durant le trajet afin de trouver comment se rendre d'un point A à un point B. De telles cartes de métro ne représentent plus une topographie métrique; elles sont plutôt conçues spécialement pour les passagers. Par exemple, la distance entre les stations est toujours la même et, dès lors, pas d'échelle. Les lignes de connexion entre les stations sont toujours dessinées en ligne droite. Parfois, le caractère construit de telles représentations laisse la place à des conclusions erronées pour l'observateur. En l'occurrence, beaucoup d'utilisateurs du métro newyorkais supposent que Manhattan est beaucoup plus grand que Brooklyn, alors qu'en réalité Brooklyn fait trois fois la taille de Manhattan.

Ces exemples des cartes de métro montrent à quel point les visualisations intégrant de nombreuses et complexes informations peuvent être plus efficaces que l'écrit en terme de séquence et d'ensemble de caractères. Toutefois, un pré-requis pour une bonne représentation visuelle est la sélection des moyens adéquats pour celle-ci, tels que les couleurs, les symboles, etc. Des résultats en psychologie cognitive montrent que les humains emmagasinent des savoirs complexes sous forme de modèles mentaux. C'est pour cela que depuis le début, nous comprenons des modèles tels qu'une carte de métro mieux qu'un texte.

Visualisation et analyse qualitative des données assistée par ordinateur (CAQDAS)

Formes de visualisation

Dans la recherche quantitative, il est habituel de présenter les résultats de simples procédures statistiques tels que des graphes par exemple, sous forme de tartes ou de diagrammes ou d'histogrammes. Les formes de visualisation qui représentent des résultats complexes et des connexions sont utilisées plus rarement. Les exemples incluent les diagrammes de trajectoire ou des modèles de causalité linéaire, dans lesquels les connexions entre un grand nombre de variables – latentes ou manifestes – peuvent être représentées.

Au sein de l'analyse qualitative de données traditionnelle, nous trouvons encore moins de visualisation que dans l'approche quantitative. C'est dans la forme d'analyse assistée par ordinateur (souvent connue sous le terme CAQDAS) que nous commençons à percevoir un intérêt pour la question de la visualisation. Nous y trouvons quatre types différents de visualisation :

- Visualisations avec le logiciel QDA lui-même. Leur principale fonction est d'améliorer la prise en charge de données durant le processus d'analyse et de rendre les données et les résultats de l'analyse plus accessibles.
- Les cartes conceptuelles. Tout comme les diagrammes de structure linéaire dans la recherche quantitative, elles permettent d'illustrer les connexions entre des concepts, des catégories, des mémos, ainsi que les interconnexions et les relations causales.
- Les visualisations interne à un cas (*within-case display*). De telles représentations présentent des séquences chronologiques (comme par exemple dans une recherche biographique) ou illustrent l'ordre et l'interconnexion des codes et des concepts. Le terme « cas » peut se référer à une variété de choses selon le matériau empirique disponible, comme par exemple une personne, une institution ou une organisation.
- Les visualisations multi-cas (*cross-case display*). L'ordre des codes, leurs chevauchements ainsi que la fréquence peuvent être représentés afin de comparer les différents cas.

Dans le cadre de cet article, les deux premières formes de visualisation ne seront pas plus explicitées. Il est possible de s'en passer car d'une part le type a nécessite des explications très spécifiques au logiciel utilisé et le type b est déjà plus connu dans la littérature sur les cartes cognitives. J'ai adopté la différence entre visualisation propre à un cas ou multi-cas d'Huberman et Miles dans leur ouvrage du milieu des années 90 « *Qualitative Data Analysis: An Expanded Sourcebook* » C'était un travail novateur dans le champ de la visualisation des données. Tout comme pour n'importe quelle visualisation, celle intervenant dans le contexte de la recherche qualitative dépend de l'acquisition de moyens efficaces. Deux questions doivent être posées :

Premièrement, quelles sont les informations pertinentes pour une analyse qualitative de données?

Dans les approches de recherche qui suivent la méthode comparative, telle que la théorisation enracinée (*Grounded Theory*), les informations pertinentes incluent les concepts théoriques, les codes et sous-codes et leurs dimensions, autant que l'ordre et la progression des codes, leur apparition simultanée ou leur proximité.

Deuxièmement, quels sont les moyens visuels disponibles pour représenter ces informations pertinentes?

Quelques exemples comprennent les symboles, les couleurs, les lignes directionnelles ou non, etc.

Les visualisations internes à un cas (Within-Case Displays)

Les visualisations internes à un cas présentent des éléments d'un cas intéressant sous forme de tableau, de graphe ou de diagramme. Un cas peut, tel qu'indiqué ci-dessus, impliquer un individu, une organisation, une institution ou une configuration. L'analyse qualitative de données comprend habituellement un texte, tel qu'une interview. Deux exemples de visualisations sont présentées dans les sections suivantes :

- *Codeliner* produit une sorte de partition du texte sélectionné, dans laquelle les codes apparaissent en tant que lignes et les parties codées sont illustrées comme des notes, apparaissant dans la chronologie de l'interview.
- *Text Portrait* crée une image qui est contruite comme une photographie. Les parties codées du texte sont présentées ligne par ligne.

Codeliner

Codeliner est une fonctionnalité utilisée pour la visualisation qui présente le texte sous forme d'une image séquentielle de son codage (Figure 1). L'axe des x reprend les paragraphes du texte. Il commence à gauche avec le numéro 1 et termine avec le dernier paragraphe du texte. L'axe des y est construit sur base des codes. La cellule du graphe indique si le code ou sous-code est attribué à un paragraphe donné ou à une section du texte. Une couleur est assignée au code spécifique et reprise dans le tableau.

Codeliner peut être mis en œuvre afin d'obtenir un aperçu rapide des sujets principaux du texte. Grâce au caractère interactif du graphe, chaque passage est accessible immédiatement par un simple click. Une telle représentation est particulièrement utile pour des analyses exploratoires, car les chevauchements et proximités peuvent être identifiées et examinées de façon approfondie.

Codeliner est particulièrement utile pour l'évaluation de *focus group*. Dans ce cas, chaque interlocuteur est représenté comme un code pour leur attribuer leur prise de parole respective. La progression des orateurs est facile à reconnaître sur le graphe. De plus, il est possible d'explorer les données pour identifier quel orateur s'engage (ou reste silencieux) sur un certain sujet. Enfin,

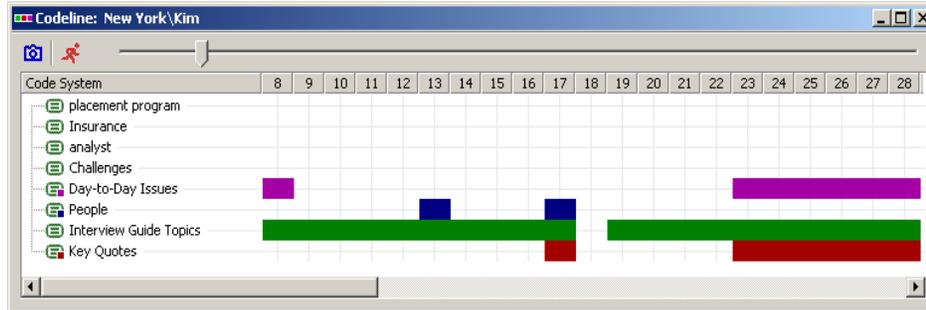


Figure 1. Visualisation des codages d'un texte

il est facile de reconnaître les effets de tels engagements sur la modération si, par exemple, le sujet ou l'orateur se modifie après l'intervention.

Text Portrait

Le second type de visualisation interne à un cas, le *Text portrait*, est une nouvelle méthode de visualisation du contenu des textes et est conçu pour attribuer des catégories (codes) aux passages du texte. Chaque catégorie est associée à une couleur. Le contenu du document est présenté de telle manière que l'attribut coloré peut être observé dans le diagramme dans l'ordre séquentiel. Le diagramme contient un nombre spécifique de carreaux colorés qui sont répartis en lignes et en colonnes.

En commençant par la première ligne et le carreau à l'extrême gauche, les carreaux sont décrits par couleurs et se suivent de gauche à droite. À la fin d'une ligne, il continue à la ligne suivante de la même façon – à la façon d'une machine à écrire. Dans une présentation normale, le nombre total de carreaux est réparti selon les passages codés du texte afin que le nombre de carreaux symbolisant un segment ou une couleur corresponde au pourcentage du segment, déterminé par la taille totale de l'ensemble des segments codés. Par exemple, si il y a seulement un segment codé, auquel deux codes sont assignés, rouge et vert, alors, dans un diagramme qui consiste en 30x40, c'est à dire 1200 carreaux, il y aura 600 rouges et 600 verts. S'il y a 3 codes (rouge, vert et magenta) pour le même segment, le diagramme contiendra 400 carreaux de chaque couleur.

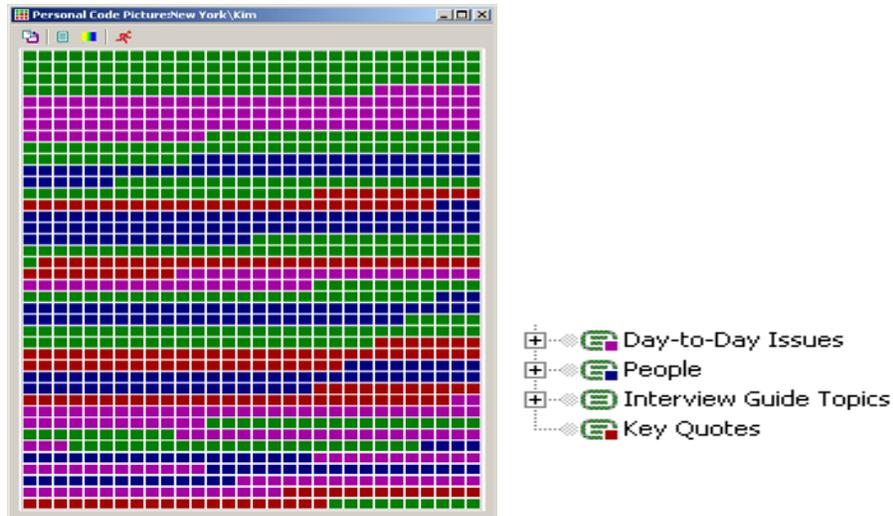


Figure 2. Exemple d'un *Text Portrait*

Text Portrait peut rendre la structure thématique d'un texte accessible sous forme d'image. Par exemple, si des couleurs sont attribuées à des émotions spécifiques dans le cadre d'une interview psychologique, l'humeur dominante ainsi que les sections particulièrement illustratives du texte peuvent être repérées d'un regard.

Les visualisations multi-cas (Cross-Case Displays)

La seconde sorte de visualisation considérée dans cet article est la visualisation multi-cas. Cette nouvelle possibilité permet de présenter des échantillons et des structures contenus dans le texte d'une façon compréhensible et de comparer les différents cas entre eux. Nous présentons deux exemples : le *Code Matrix Browser* et le *Code Relation Browser* de MAXQDA. Les deux représentations visuelles permettent de présenter les codages d'une façon graphique. Avec un coup d'œil à la matrice, on peut extraire quel texte appartient à quelle catégorie et le nombre (élevé ou non) de segments qu'elle contient. Cette visualisation est plus vérifiable et interprétable qu'une matrice numérique.

L'organigramme de la Figure 3, basé sur les travaux de Nadin (1998), illustre la progression des données à la reconnaissance des régularités et des structures :

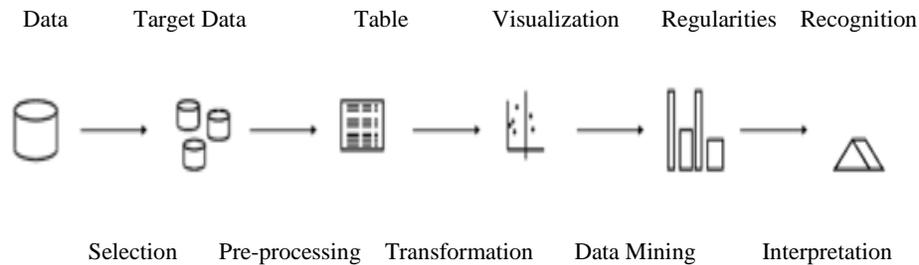


Figure 3. La progression des données à la reconnaissance des structures

Cela signifie qu'une présentation multi-cas des codes commence par la sélection des textes et des codes et sous-codes à visualiser. L'étape décisive est la transformation d'un tableau en graphe, en ce que les régularités sont plus facilement reconnaissables dans un graphe que dans un tableau qui contient une grande quantité d'information. Par exemple, afin de pouvoir identifier le plus grand nombre dans un tableau, il faut lire et comparer tous les nombres entre eux. Mais avec une visualisation où le nombre le plus élevé est présenté sous une forme ou une couleur spécifique, une reconnaissance immédiate est possible.

Afin de pouvoir comparer des textes à partir de leur codage, une visualisation croisée peut être activée. Les textes sont présentés les uns à côté des autres dans la présentation pour faciliter leur comparaison directe au regard des catégories. Le *Code Matrix Browser* nous donne un diagramme tel que le montre la Figure 4.

Les nœuds individuels dans la matrice symbolisent – à travers leur taille et couleur – combien de codages du code ou sous-code spécifique sont issus du texte. Les colonnes de la matrice sont construites à partir du texte : au plus le nœud est grand dans la colonne, au plus on peut trouver de segments codés dans le texte qui entrent dans cette catégorie ou sous-catégorie. Avec un double « clic » sur le nœud, tous les segments du texte correspondants sont listés. La matrice ne présente donc pas seulement une vision globale du codage mais contient également – même de façon invisible – les différents morceaux de texte qui se retrouvent derrière les nœuds.

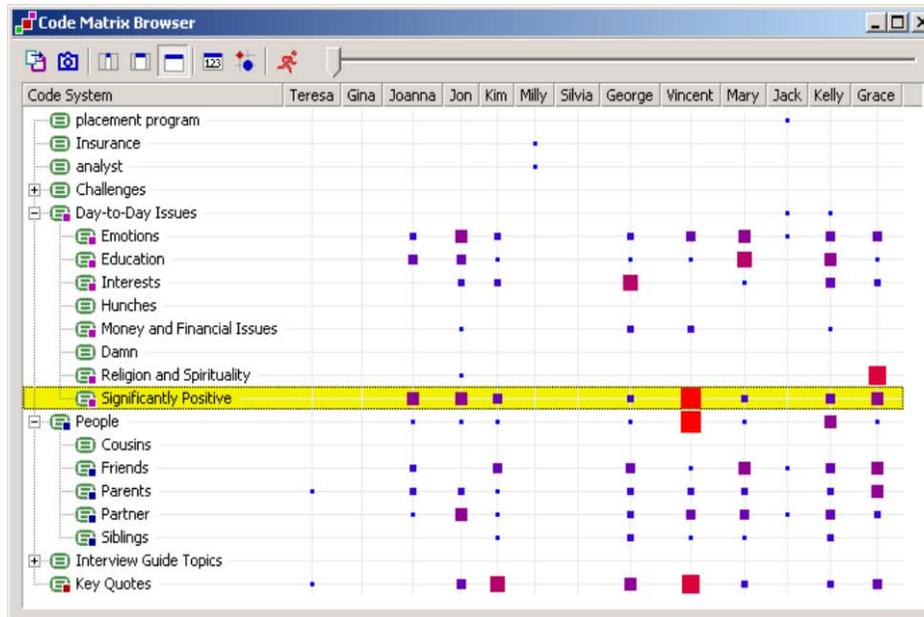


Figure 4. Diagramme des codages dans les textes

La seconde forme de visualisation multi-cas présentée ici affiche les connexions entre les codes, par exemple, combien de fois les codes se chevauchent ou sont attribués simultanément à un même passage de texte. Localiser les chevauchements entre catégories manuellement est une tâche trop énergivore, alors qu'elle peut être réalisée rapidement et de façon fiable à l'aide de l'ordinateur. La Figure 5 montre une présentation du *Code-Relations Browser* dans MAXQDA. Les chevauchements des différents codes sont analysés grâce à cette matrice symétrique des codes. À la place du nombre de chevauchements trouvés, les nœuds de tailles et de couleurs différentes apparaissent dans les cellules de la matrice : au plus le nœud est large, au plus nombreux sont les chevauchements trouvés dans le matériau. L'information numérique apparaît dès que le curseur de la souris est positionné sur le nœud. Un double-clic permet d'extraire du texte l'ensemble des passages correspondant.



Figure 5. Diagramme de chevauchements des codages

Conclusion

Dans la recherche sociale empirique, nous devrions parvenir à travailler avec des visualisations, tout comme d'autres disciplines scientifiques reconnues ont pu les pratiquer avec succès. Toutefois, nous devons apporter une attention particulière à l'accroissement de notre propre compréhension et à l'amélioration de la présentation extérieure de nos résultats. Le proverbe approprié de Confucius « Une image vaut mille mots » est particulièrement valable ici tant la publicité donnée par les images dévoile bien plus que les milliers de mots de scientifiques en sciences sociales pour expliquer le contenu. Les images ont acquis un rôle prédominant dans notre société; elles nous entourent et laissent des impressions durables dans notre mémoire – souvenons-nous du 11 septembre.

Les visualisations deviendront plus complexes et compliquées dans le futur. Si, par exemple, un médecin doit évaluer une radiographie ou une raisonance magnétique, il doit apprendre à le faire. Nous, en tant que profane, nous ne reconnaissons pratiquement rien lorsque nous regardons de telles images. Mais si quelqu'un nous les explique, par exemple une échographie lors d'un diagnostic prénatal de grossesse, alors nous apprenons rapidement et sommes capables de reconnaître la tête, les bras, les jambes et même les ventricules du cœur.

C'est l'approche nécessaire pour rendre observable et compréhensible les analyses visuelles de données et de problèmes en sciences sociales.

Références

- Banks, M. (2001). *Visual methods in social research*. London : Sage.
- Blasius, J., & Greenacre, M.H. (1998). *Visualization of categorical data*. San Diego : Academic Press.
- Feldmann, K. (2003a). Du sollst dir kein Bild machen! (Nicht)Visualisierung in der Soziologie. *TRANS. Internet-Zeitschrift für Kulturwissenschaften*, 14. Document consulté le 15 juillet 2009 de <http://www.inst.at/trans/14Nr/feldmann14.htm>.
- Huberman, A.M., & Miles, M.B. (1994). Data management and analysis methods. Dans N. Denzin, & Y. Lincoln (Éds), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA. : Sage.
- Krempel, L. (2005). *Visualisierung komplexer strukturen. Grundlagen der darstellung mehrdimensionaler Netzwerke*. Frankfurt, M. : Campus.
- Nadin, M. (1998, 27 Novembre). *Zwischen daten und display. Communication présentée au Third Visualization Workshop, University of Paderborn*. Document consulté le 15 juillet 2009 de <http://www.nadin.name/pdf/between.pdf>.

Udo Kuckartz est professeur au département d'éducation de l'Université de Marburg (Allemagne). Il est titulaire d'un doctorat en sociologie et d'un master en sociologie et science politique. Ses domaines d'intérêt se situent en sociologie de l'environnement, en recherche qualitative et en méthodes mixtes. Il est l'auteur du logiciel MAXQDA qui est utilisé dans plus de 70 pays.