



**HAL**  
open science

## Organisation des connaissances, données et algorithmes : place aux visualisations !

Viviane Clavier

► **To cite this version:**

Viviane Clavier. Organisation des connaissances, données et algorithmes : place aux visualisations !. 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?, Oct 2019, Montpellier, France. hal-02306622

**HAL Id: hal-02306622**

**<https://hal.science/hal-02306622>**

Submitted on 6 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.

# Organisation des connaissances, données et algorithmes : place aux visualisations !

## **Viviane Clavier**

MCF HDR en Sciences de l'information et de la communication

Gresec, Université Grenoble Alpes

[viviane.clavier@univ-grenoble-alpes.fr](mailto:viviane.clavier@univ-grenoble-alpes.fr)

## Résumé

L'article montre comment les visualisations ont progressivement pris le pas sur les modèles de représentation des connaissances classiques de nature logocentrée. Se situant dans une approche théorique et historique, cet article s'attache aux conséquences des changements d'échelle liés à la production de données sur les modèles d'organisation des connaissances. Jusqu'alors confinés à un cercle étroit de chercheurs spécialisés en traitement de données textuelles et traitement automatique des langues, les outils de datavisualisation sont désormais accessibles au plus grand nombre. Ce faisant, de nouveaux enjeux se profilent pour l'organisation des connaissances, de nouveaux métiers et institutions de recueil de données émergent qui contribuent à redéfinir les normes, les standards et les modes de représentation des connaissances.

## Mots clés

Organisation des connaissances, visualisations, langages, data, algorithmes.

## Title

Knowledge Organization, Data and Algorithms : the new Era of visual Representations

## Abstract

This article shows how visual representations have progressively taken the lead over classical language-based models of knowledge organization (KO). The paper adopts a theoretical and historical perspective and focuses on the consequences of the changes in the volume of data generated by data production on the KO models. Until now, data visualization tools have been used mainly by researchers with expertise in textual data processing or in computational linguistics. But now, these tools are accessible to a greater number of users. Thus, there are new issues at stake for KO, other professions and institutions for gathering data which contribute to defining new standards and KO representations.

## Keywords

Knowledge Organization, visual Representations, Languages, Data, Algorithms.

## INTRODUCTION

Notre contribution s'attache à la place des langages dans le processus de collecte, de représentation et d'organisation des connaissances dans un contexte marqué par la production massive de données. En suggérant dans le titre que les visualisations s'imposent pour représenter les connaissances, nous signifions implicitement qu'elles ont supplanté les langages. Nous ne disons pas qu'ils ont disparu, l'usage des langages contrôlés étant toujours d'actualité pour la bibliothéconomie dans les institutions patrimoniales et dans plusieurs champs de l'information spécialisée. Cependant, il nous semble que l'attention portée aux langages s'est détournée vers d'autres objets de connaissance (les algorithmes), d'autres sémiotiques (la visualisation de données), d'autres niveaux de représentation (les données textuelles), d'autres finalités de l'organisation des connaissances (la navigation plutôt que la catégorisation), d'autres propriétés (la transparence et la simplicité plutôt que la précision et l'exhaustivité). Le dernier rapport 2018 de l'IFLA [1] sur les tendances de l'information et la manière dont elle façonne les bibliothèques est d'ailleurs révélateur de cette situation puisque le terme de « langage » n'est pas mentionné.

Ainsi, cette communication pose un ensemble de questions, émet diverses hypothèses et s'interroge sur les raisons qui expliqueraient la remise en question des langages, et s'ils sont remplacés, par quoi ils le sont et comment. La présente étude comporte une dimension diachronique plutôt qu'historique perceptible à travers les travaux de recherche en sciences de l'information et la communication (CPDirSIC, 2018) et fait également référence à notre propre recherche dans laquelle le langage, les textes et les discours ont toujours joué un rôle central (Clavier, 2014). Notre présentation comporte deux parties : la première montre comment les représentations visuelles ont progressivement pris le pas sur les représentations logocentrées dans les systèmes d'organisation des connaissances. La seconde partie montre comment cette tendance observée dans les travaux de recherche, s'est également répandue auprès d'acteurs de la société civile ou de professionnels concernés par l'information et la communication. Il ressort cependant que les langages restent bien présents, même s'ils apparaissent relégués à une fonction descriptive de second plan.

## 1 – ÉVOLUTION ET PERMANENCE DES ENJEUX DE L'ORGANISATION DES CONNAISSANCES

### 1.1 Perspective appliquée : les langages pour les systèmes d'organisation des connaissances

Nous restreindrons la notion de « langages » aux systèmes symboliques choisis pour coder des unités de langue et des significations et qui sont utilisés pour alimenter des systèmes d'organisation des connaissances (SOC), « *des outils destinés à ordonner les ressources de savoir enregistrés* » (Sosińska-Kalata, 2012). Les SOC désignent « *toutes sortes de schémas d'organisation allant des simples listes alphabétiques ou faiblement structurées (listes d'autorité, glossaires, dictionnaires, nomenclatures, etc.) à des schémas classificatoires hiérarchiques (plans de classement, classifications générales ou spécialisées, taxinomies, listes de vedettes matières, etc.) ou encore à des organisations privilégiant des relations non exclusivement hiérarchiques (thésaurus, réseaux sémantiques, ontologies, etc.)* » (Polity et al., 2005, p. 13). Cette perspective nous place d'emblée dans ce qu'il convient d'appeler *le sens étroit* de l'organisation des connaissances qui renvoie à la confection de produits d'information destinés à classer, organiser et structurer des connaissances pour faciliter l'accès à l'information [2].

Le point commun entre les langages documentaires, les classifications, les folksonomies, les ontologies, etc. et aujourd'hui les *datas* réside dans la mise en œuvre de méthodes de traitements pour obtenir une représentation de contenus. Ainsi, quelle que soit l'époque, les connaissances sont le résultat de choix et de transformations qui président à leur sélection, le cas échéant, leur normalisation et leur catégorisation. Le caractère massif des données à traiter n'est pas une nouveauté, le développement de l'informatique documentaire reposait dès les années 70 sur l'idée que l'automatisation permettait de réduire les coûts d'indexation manuelle d'une documentation en constante augmentation à laquelle les professionnels ne pouvaient faire face (Chaumier, 1982). Ce qui nous interroge ici concerne la nature des connaissances sélectionnées en regard des choix de traitements informatiques.

## 1.2 Des années 80 à aujourd'hui : la fin des langages ?

Entre les années 80-90, avec l'essor des systèmes de recherche d'information en full text et le développement d'applications comme l'extraction de connaissances soutenues par des technologies linguistiques (Chaudiron, 2007), les langages étaient analysés à l'aune de la linguistique (Van Slype, 1982). La rencontre entre la linguistique automatique et la documentation (Rouault, 1987) constituait un lieu de rapprochement entre « *ces deux disciplines alors même qu'un mur les sépare* » (Van Slype, 1982, p. 87). Conçus dans une perspective logico-syntaxique des connaissances manipulables par des programmes informatiques, les langages naturels ont fait l'objet de débats théoriques et d'enjeux socio-économiques forts. Les questions qui se posaient étaient les suivantes (Lallich-Boidin et Maret, 2005) : Quelle est l'unité de signification pertinente : le mot, le morphème ? Quels avantages et inconvénients présentent les langages libres par rapport aux langages contrôlés ? Comment représenter la signification finement en tenant compte des ambiguïtés, de l'implicite ? Faut-il recourir à des ressources extérieures pour définir des langages d'indexation (thésaurus, vocabulaires contrôlés) ou aux documents eux-mêmes, cette réflexion mettant en tension la fidélité de l'indexation en regard de l'économie de coût pour maintenir des ressources lexicales. Selon cette conception, les langages sont travaillés avec et par les linguistes « talistes » et sont interrogés dans leur rapport aux unités de langue. Cette approche centrée sur une description fine du langage, reposant sur une approche compositionnelle du sens dirigé par la syntaxe sous forme de représentations logiques, se voit supplantée par une approche fondée sur une représentation empirique du sens dans laquelle on ne cherche plus à comprendre la langue mais à décrire des corpus textuels.

Ainsi, les années 90 constituent-elles un tournant marqué par le développement du Web, laissant place à la linguistique de corpus (Habert et al., 1997), une approche qui déplace les questionnements vers la constitution de données textuelles et leur description. S'ouvre la période des « gros corpus » destinés à l'annotation par des outils d'exploration permettant d'apprécier la représentativité des phénomènes langagiers. Selon cette perspective, le caractère « vaste » des données est encore mis en avant et l'annotation de textes, le tagging (ou étiquetage), vise le recueil de typologies fondées sur des corrélations de traits linguistiques (Biber, 1993). Ces méthodes d'annotation, souvent manuelles ou semi-automatiques, couplées à des méthodes de traitement de l'information plus robustes, comme la fouille de textes fondée sur l'apprentissage, visent des tâches de classification automatique supervisée ou non supervisée. Les connaissances issues de l'annotation de corpus peuvent être des catégories morpho-syntaxiques (Poudat et al., 2006), des collocations (Tutin, 1997), la phraséologie ayant bénéficié du développement conjoint des méthodes symboliques et des méthodes probabilistes par l'apprentissage automatique. Ainsi, contrairement aux connaissances produites par le TAL qui visent l'automatisation d'applications, les

connaissances qui résultent des annotations produites par la linguistique de corpus « *repose[nt] sur une itération entre l'analyse des sorties logicielles et la consultation « humaine » des textes ou de fragments de textes, par exemple les concordances* » (Valette et Egle, 2014). La production de connaissances s'inscrit alors dans une approche outillée destinée à explorer les données afin de faire émerger des récurrences linguistiques (des expressions figées, des locutions, etc.). Cette perspective a conduit au développement de méthodes fondées sur des approches linguistiques « opportunistes » i.e. qui exploitent des phénomènes de surface identifiables automatiquement (*ibid.*, p. 110).

Avec le temps, la démarche empirique s'est installée, et l'engouement pour les méthodes de traitement numérique s'est renforcé. Le développement d'outils d'analyse de données ainsi que la généralisation de méthodes destinées aux sciences sociales et au « grand public » se sont accompagnés, comme l'indique Dominique Boullier (Boullier, 2015), de « la croyance populaire qui annonce la fin de la science et des théories scientifiques ». Ainsi s'est ouverte une nouvelle période dans laquelle les Big Data « *seraient des mesures effectives du réel* » (*ibid.*). Selon cette perspective, les données s'appuient sur le recueil de sacs de mots issus du Web qui se caractérisent par leur volumétrie et que l'on peut visualiser dans des espaces à deux dimensions. Il s'agit alors d'interpréter et de découvrir ces connaissances dont le statut n'est pas toujours très clair : les connaissances représentent-elles des thèmes, des sujets, des mots-clés, des discours, des univers lexicométriques, etc. Le point sombre de ces outils réside souvent dans l'opacité qui entoure les étapes de prétraitement des données, le choix des catégories (lexicales, textuelles) ou d'unités de discours.

Comme l'indiquait Pascale Sébillot en 2002, les méthodes mixtes se sont certes développées, mais dans des domaines circonscrits et avec un objectif qui n'est plus de comprendre finement les textes mais d'obtenir des représentations de sens utiles pour des applications précises (Sébillot, 2002). Ainsi, les systèmes à base de règles sont utilisés afin de développer des ressources pour l'analyse des sentiments et de l'opinion (Poibeau, 2014), ce qui permet d'annoter des données destinées à entraîner des classifieurs, d'autres applications s'appuient également sur la profusion de données comme la traduction automatique, le journalisme de données ou la vérification par les faits. Cependant, comme l'indique Thierry Poibeau (*ibid.*), la part liée au TAL est variable, et « *les difficultés récurrentes concernent l'adaptabilité des systèmes, le temps nécessaire pour développer des ressources pour un nouveau domaine et la disponibilité ou non d'exemples en quantité suffisante* ». Le chercheur signale également que la définition des informations recherchées et la qualité des résultats obtenus sont également des questions régulièrement posées.

Depuis de nombreuses années déjà, il existe un fort engouement pour la visualisation ainsi que le révèle cette revue de littérature récente réalisée par des chercheurs de l'Université de Swansea sur les ouvrages publiés sur ce sujet (Rees et Laramee, 2019). Les auteurs font remonter les recherches à 1967 l'une des premières monographies consacrée à la Sémiologie graphique de Jacques Bertin. Selon eux, les 40 années qui ont suivi ont été peu prolifiques, mais à partir des années 2000, il s'est ensuivi une véritable frénésie de publications au point qu'il est difficile actuellement d'en suivre les parutions (*ibid.*, p. 610). La visualisation doit être comprise à la fois comme un processus de mise en visibilité et comme le résultat de ce processus. Dans le premier cas, il s'agit de « *l'aboutissement d'une instrumentation d'exploration de masses de données devenant aptes à générer des indicateurs, cartographies, etc.* » (Reymond, 2016, p. 11). Dans le second cas, les visuels produits constituent les matériaux de l'infographie dont les réagencements sont constitutifs d'une écriture à part entière. Cependant, il peut s'agir de visualiser des données, des informations ou des connaissances. La « *visualisation de données* » est le terme générique utilisé pour dénommer les outils de « *dataviz* » et consiste à « *opérer une transformation sémiotique entre les résultats*

*d'une analyse de données (numériques, catégorielles, textuelles) et une représentation graphique* » (Hachour, 2015). La « visualisation de l'information » est une discipline de l'informatique qui « utilise des représentations visuelles et interactives de données abstraites sur un ordinateur pour amplifier la cognition » (Lamy, 2017, p. 76 sqq) avec 6 objectifs possibles : afficher une grande quantité de données, faciliter la recherche d'une information donnée, détecter des motifs, permettre des inférences visuelles, surveiller la venue d'événements, explorer des jeux de données. Les représentations graphiques obtenues servent plusieurs applications destinées à visualiser l'information, comme l'information journalistique ou documentaire. Ainsi, la datavisualisation au service de l'information (Yikun et Zhao, 2016) permet de découvrir des actualités journalistiques en établissant « *des interactions nouvelles* » et de les présenter de manière visuelle. Quant à la visualisation de l'information documentaire, elles peuvent être visualisées par des cartographies destinées à naviguer dans des classifications (Dewey), ou au sein de répertoires (Rameau) ou dans un catalogue (Papy, 2005). Enfin, la « visualisation de connaissances » s'inscrit dans le champ de recherche de l'intelligence artificielle et s'appuie sur la représentation des connaissances, leur modélisation et leur visualisation. Selon le chercheur en informatique Jean-Baptiste Lamy (Lamy, 2017, p. 12), l'on distingue la visualisation iconique des connaissances qui consiste à traduire les connaissances à l'aide d'un langage iconique comprenant un lexique de pictogrammes et une grammaire ; et une visualisation structurelle des connaissances qui consiste à représenter graphiquement la structure des connaissances à l'aide de techniques de visualisation.

## **2 – DIFFUSION DES CONNAISSANCES : PRIORITÉ AUX REPRÉSENTATIONS VISUELLES POUR DE NOMBREUX ACTEURS**

Le choix des sémiotiques verbales ou visuelles pour représenter les connaissances dans les systèmes d'information ne se pose pas uniquement dans une perspective scientifique mais s'inscrit plus largement dans un contexte spatio-temporel, socio-économique et culturel. Actuellement, les visualisations s'imposent comme modes d'expression dans notre société : présence d'infographies, d'images fixes ou animées (vidéos) dans le domaine de la communication publique ou de la communication scientifique, dans les médias, sur le web. Pour lever toute ambiguïté, la question n'est pas de montrer que nous vivons dans « un monde d'images » comme l'indique le philosophe Frank Robert en introduction de l'ouvrage collectif « *Philosophies de l'image* », mais que l'importance prise des visualisations – les « visuels » comme disent les professionnels de la communication – comme formes de sémiotisation influence les modes de représentation et d'organisation des connaissances. Dans la suite, nous présentons quelques tendances observées auprès d'acteurs qui, pour une raison ou une autre, jouent un rôle important dans la diffusion des normes, des standards ou des formats de représentation des connaissances. Ces exemples – qui ne sont pas issus de la constitution d'un corpus raisonné –, illustrent le fait que les connaissances sont toujours ancrées dans les époques, les lieux et dépendent probablement des modes.

### **2.1 Promotion de l'image dans les outils et techniques de rédaction pour le web**

L'image est promue par les professionnels du web qui définissent les standards et techniques d'écriture destinés à optimiser la visibilité de sites internet via leur référencement et leur positionnement dans les moteurs de recherche. Ainsi, il existe de très nombreux sites destinés aux rédacteurs web qui émettent des préconisations en matière d'écriture sur le web, édictent des règles de concision, de simplicité, de structuration de contenu, d'adresse, etc [3].

Lorsqu'il est question d'images, il peut s'agir d'images fixes ou animées, telles les vidéos. Sur le site Annei, l'image et le texte, « *considérés de tous temps comme des rivales* » sont présentés aujourd'hui comme complémentaires moyennant quelques ajustements destinés à ne pas cantonner l'image dans un rôle secondaire de faire-valoir. Il est ainsi conseillé d'associer un « *texte alternatif* », un « *équivalent textuel, descriptif et concis* » des images qui ne dépassent pas « *250 caractères* », ou « *de prévoir une retranscription des vidéos dont la bande son sera rendue accessible par une retranscription au format texte, détectable par les moteurs de recherche* ». En matière de choix d'images, des préconisations sur le contenu sont également émises : « *Choisissez des images à valeur informative* », ce conseil vise à écarter les images à valeur illustrative. Sont ainsi considérées comme informatives : « *une infographie, un schéma, une photographie prise en situation réelle, apportant plus qu'un visuel pompé d'une banque d'images clichées à valeur uniquement décorative* » (ibid.). Dans ce dernier cas de figure, il est préconisé « *d'ajouter une légende, concise, qui donne du sens à l'image, chaque fois que c'est possible* » afin de renforcer le « *taux d'attention à l'image* ».

Lorsque les conseils rédactionnels s'appliquent à des médias sociaux, l'image devrait même remplacer le texte au motif que la présence d'une image assure une visite sous la forme d'un clic et garantit un trafic générateur de profit ou d'audience. Ainsi, cette préconisation serait-elle liée aux usages observés des utilisateurs qui cliquent sur les images pour naviguer sur internet :

*« Les utilisateurs de Pinterest ou de Facebook sont coutumiers de la navigation en zigzag. Un clic mène à un autre clic, une image mène à un lien où se trouvent d'autres images, qui elles aussi mènent vers d'autres liens. L'image, ce sont les petits cailloux du Petit Poucet. Semées savamment, elles apportent bien plus de trafic sur un site internet qu'une simple quote ».*

*« Quand on sait que les photos postées sur Facebook peuvent générer 53% plus de likes sur un post, on ne s'en prive pas. Les utilisateurs sont bien plus enclins à liker ou à suivre une marque si celle-ci est active et partage des images. Il est fini le temps de la promotion textuelle. »*

Les standards de l'écriture web sur les médias sociaux orientent désormais vers le choix d'une sémiotique visuelle : il serait préférable de « *montrer* » plutôt que de « *raconter* », les « *mots sont là pour souligner l'image* », il faut user de « *stratagèmes visuels pour conquérir le visiteur* », etc. Pour autant, la sempiternelle question de la description des documents, de leur représentation et de leur indexation reste posée dans le cadre de la recherche d'information via les requêtes des usagers. Ainsi, l'image ramène-t-elle toujours in fine à la question des langages et de leur choix pour décrire des images. Cette question est plus spécifiquement abordée sur les sites destinés à développer la visibilité de sites internet dans les moteurs de recherche (*Search Engine Optimization*) ou sur les médias sociaux (*Social Media Optimization*) ou encore spécifiquement dans les moteurs dédiés au marketing (*Search Engine Marketing*). Il est fait référence au W3C qui définit les langages et standards du web et énonce les formats d'images pris en charge par les moteurs de recherche, les balises de description et les conseils pour décrire les attributs alt et title qui permettent de décrire « *le contenu d'une image de façon claire et concise* ».

*« Le texte alternatif des images : Google référence aussi les images ! Pensez à toujours renseigner le texte alternatif (balise alt) de vos images d'articles. Introduisez-y votre mot-clé principal au sein d'une expression courte. La balise alt régulièrement omise est celle de l'image à la une, ce qui a pour effet d'amoindrir le référencement de ses images. Pour y remédier, il suffit de renseigner le champ Texte alternatif au moment de définir votre image à la une. » (ibid.)*

Comme on peut le constater, l'image, loin de supplanter le texte, pose à nouveau la question de la représentation textuelle, du choix des mots et de leur vertu descriptive, renversant les rôles de l'image « illustrative » au service du texte. Du côté de la recherche d'information, ces tendances ont conduit à développer des recherches sur la classification automatique d'images, la reconnaissance faciale, l'élaboration de langages de description et d'annotation des images, telles les ontologies. Le moteur de recherche « Google Image » connaît d'ailleurs un succès grandissant auprès des utilisateurs au point que le référencement naturel d'images apparaît désormais comme une priorité [4].

## 2.2 Développement d'outils de visualisation de données pour le grand public

Jusqu'alors réservées à des spécialistes en statistiques descriptives multidimensionnelles, certaines méthodes comme l'analyse en composantes principales (ACP) et l'analyse factorielle des correspondances (AFC) ont beaucoup été utilisées en SHS pour étudier des textes littéraires ou politiques dans différentes dimensions : le lexique, le vocabulaire, le style, les thèmes, etc. Une littérature importante est consacrée aux statistiques textuelles depuis la fin des années 50 et dont le point commun réside dans la production de systèmes d'organisation de connaissances sous formes de sémiotiques verbales : des dictionnaires (Guiraud, 1959), des index, des concordances, des segments répétés, des séries textuelles chronologiques, des corpus parallèles, etc. (Salem et Lebart, 1996). Dans ces travaux, les visualisations – des tableaux lexicaux entiers, des figures représentant des plans factoriels, etc. – ne sont pas une finalité mais elles sont pensées comme une « aide à la lecture d'une série de textes » (*ibid.* p. 135sq). Ces ouvrages d'introduction donnent d'ailleurs des clés de lecture : « Comment interpréter les distances ? », « Exemples d'applications », « Lecture de la figure », etc. Il apparaît ainsi que l'interprétation des données textuelles est d'abord soumise à la compréhension des réorganisations lexicométriques présentées sous formes de visualisations, lesquelles sont le résultat d'un ensemble de traitements et de choix sur la taille des corpus textuels, leur représentativité, leur segmentation, la définition des unités de décompte, etc. Dans la continuité de ces méthodes, la fouille de texte intègre les traitements statistiques de données mais aussi d'autres tâches, comme la classification supervisée ou non supervisée (Ibekwe-SanJuan, 2007).

Développées par des informaticiens spécialisés en apprentissage, ces méthodes abordent les textes comme des données destinées à entraîner des classificateurs, dont les scores sont mesurés et comparés. Les applications comme la recherche d'information, la détection de motifs, la veille scientifique, etc. sont toujours envisagées, si bien que le texte, au sens fort du terme et abordé dans ses conditions de production, n'est pas un objet d'attention. Il nous semble que c'est ce glissement d'une conception sacralisée du texte vers une vision instrumentale comme corpus d'entraînement qui est à l'origine d'une interprétation centrée sur la compréhension des tendances, des régularités, des effets massifs... et moins sur ce qui est de l'ordre de l'invisible que seule la connaissance des textes permet de restituer.

Dans le même temps, les outils de dataviz se démocratisent, de nombreux acteurs étant en demande d'analyses de discours sur les réseaux sociaux, les blogs, Twitter, les forums, etc. Ainsi, pour des raisons économiques, les instituts de sondage ont introduit des méthodes dites passives inspirées du Big Data, qui sont qualitatives et automatisables et utilisées en complément d'enquêtes quantitatives. La demande sociale est large : marketing, sondages politiques, consommation, etc. Les sites web qui répertorient les outils de dataviz sont légion et distinguent différentes familles : les cartes mentales, la visualisation de réseaux de données relationnelles, la cartographie de l'information, etc. Leur succès tiendrait à leur qualité de transparence, de simplicité et de fidélité aux données. En atteste cet exemple de discours [5]



d'annonce présentant les atouts de la dataviz pour représenter des données textuelles ou numériques.

*« La dataviz, ou la visualisation de données, est une pratique que l'on côtoie au quotidien sans se rendre compte, ne serait-ce qu'en ouvrant un journal ou en regardant la télévision. L'exemple le plus simple reste le sondage. Avec l'ère numérique, elle est devenue un puissant outil de communication. »*

### ***Dataviz : une définition simple à comprendre***

*Dans une société de plus en plus attirée par l'aspect graphique, la visualisation de données prime sur la donnée brute. Elle aide à éclairer des informations en apparence complexes ou noyées dans une grande quantité de paramètres. Le terme dataviz désigne donc l'ensemble des représentations visuelles de ces données brutes. Le but premier ? Éclaircir un ou des phénomènes en lui donnant un aspect plus ergonomique qu'un tableur rempli de chiffres. »*

La transparence fait référence au caractère immédiatement perceptible et compréhensible du message dès lors qu'il est représenté sous une forme visuelle, ce qui lui conférerait des vertus communicationnelles incontestables. La simplicité renvoie à la capacité des outils à représenter de manière synthétique sous une forme visuelle des informations exprimées sur un mode verbal, jugé plus complexe. Enfin, la fidélité résulte de la capacité des outils à représenter les données telles qu'elles émanent des individus sans aucune forme de médiation. Ces propriétés présentées comme des atouts techniques s'inscrivent dans la continuité de débats théoriques portant sur les différences de statut sémiotique du signe pictural et du signe scriptural (Christin, 2012). Ainsi, ces discours renouent-ils avec les débats stériles sur le caractère plus transparent du signe pictural face au signe scriptural au motif que l'interprétation dispense de passer par un code arbitraire pour interpréter une image contrairement à un texte ; de même la simplicité fait sans doute référence au faible nombre de signes en jeu, un ensemble réduit de formes visuelle étant toujours plus synthétique qu'une infinité de signes verbaux combinés pour décrire une langue. Des débats loin d'être clos et qui interrogent la pertinence de tel ou tel processus d'encodage des connaissances sans prise en compte des conditions de réception. Enfin, la fidélité est également une question largement débattue en information-documentation au sujet des langages d'indexation libres face aux langages contrôlés.

La représentation de connaissances sous forme visuelle est également investie par les professionnels de la communication qui abordent le design de l'information sous un angle esthétique. De nouvelles compétences sont attendues pour les professionnels de l'information communication en matière de dataviz, qui se trouvent à la croisée de l'informatique, des statistiques et de la communication (Arruabarrena, 2017). Enfin, la visualisation des connaissances est érigée au statut d'objet artistique digne d'être exposé. Ainsi, l'exposition temporaire au Mundaneum de Mons intitulée « Mapping Knoweldge » présente les travaux de Paul Otlet, plus connu pour la création de la classification décimale universelle (CDU), que pour ses textes et dessins destinés à « synthétiser ce qui est connu » (cf. Figure 1).



Figure 1. Capture d'écran de l'exposition consacrée aux visualisations de Paul Otlet à partir de la plateforme GoogleArts&Culture

## CONCLUSION

S'intéresser à la place des visualisations comme systèmes d'organisation des connaissances s'inscrit dans une longue filiation de recherches qui discutent les choix de vocabulaires, leur mode d'organisation et de structuration pour accéder aux savoirs. Cette contribution illustre que les visualisations obtenues par traitements de données, ne sont ni « naturelles », ni « simples », ni « objectives » bien qu'elles soient réalisées par des outils automatiques et qu'elles s'appliquent à des données qui sont déjà-là (Bullich et Clavier, 2018). Ce constat a notamment été émis par Michèle Hudon et Widad Mustapha El Hadi à propos des classifications : « Il n'existe pas de structure d'organisation naturelle et toute structure classificatoire ne peut être qu'inventée et construite, jamais découverte. De plus, l'établissement d'un cadre ou d'une structure classificatoire témoigne toujours de la rencontre et du choc d'impératifs philosophiques, technologiques, sociaux, économiques et politiques (Svenonius, 1991) » (Hudon et Mustafa El Hadi, 2010, p. 12).

Portée par des disciplines spécialisées dans le traitement de données massives et encouragée par les politiques de recherche et d'éducation, la visualisation de données est, pour les professionnels de la gestion des connaissances, plus qu'un outil, il s'agit véritablement d'une aide à la découverte de connaissances (Moreux, 2016). Ainsi, il ne s'agirait plus de nommer des connaissances pour les comprendre, mais de les montrer pour les rendre visibles, une perspective qui laisse toute sa place à de nouveaux acteurs, de nouveaux métiers, de nouvelles structures de recueil de données qui contribuent à redéfinir les normes, les standards, ainsi que les formes d'écritures et les modes de représentation des connaissances.

## NOTES

[1] IFLA, *A la crête des vagues, ou submergée par la marée ? Naviguer dans un environnement de l'information en mutation. Aperçu du rapport sur les tendances*, 2018, 17p. Disponible sur : [https://trends.ifla.org/files/trends/assets/ifla-trend-report\\_french.pdf](https://trends.ifla.org/files/trends/assets/ifla-trend-report_french.pdf) (consulté le 30 décembre 2018).

[2] Cette perspective se situe entre une conception *large* et théorique des connaissances – pour une conception ontologique des connaissances (Gnoli, 2008) ou sociales des connaissances (Hjørland, 2015) – et une approche *intermédiaire* centrée sur les applications

et les formes de médiation des connaissances observables en situation professionnelle (Clavier et Paganelli, 2013 ; 2015).

[3] Les exemples de sites mentionnés figurent parmi les premiers résultats de Google à la requête posée « rédaction web image et texte ».

[4] [https://www.1ere-position.fr/blog/optimiser-le-referencement-dimage-sur-google-images/#Google Images en forte croissance](https://www.1ere-position.fr/blog/optimiser-le-referencement-dimage-sur-google-images/#Google%20Images%20en%20forte%20croissance) (consulté le 2 mai 2019)

[5] <https://www.lebigdata.fr/dataviz-qu-est-ce-que-c-est>

## RÉFÉRENCES BIBLIOGRAPHIQUES

ARRUABARRENA Béa (2017). L'expert en dataviz, un métier en transition. *I2D – Information, données & documents*, 2017, vol. 54, n° 3, p. 7-8.

BIBER Douglas (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, 1993, vol. 19, n°2, p. 243-258.

BOULLIER Dominique (2015). Les sciences sociales face aux traces du *big data*. Société, opinion ou vibrations ? *Revue française de science politique*, 2015, vol. 65, no. 5, p. 805-828.

BULLICH Vincent et CLAVIER Viviane (2018). Dossier 2018. Production des données, "production de la société". Les Big Data et algorithmes au regard des sciences de l'information et la communication. *Les Enjeux de l'information et la communication*, 177p. Disponible sur : <https://lesenjeux.univ-grenoble-alpes.fr/pageshtml/art2018.html#dossier> (consultée le 15 mai 2019).

CHAUDIRON Stéphane (2007). Technologies linguistiques et modes de représentation de l'information textuelle. *Documentaliste-Sciences de l'Information*, 2007, vol. 44, n°1, p. 30-39.

CHAUMIER Jacques (1982). *Analyse et langages documentaires. Le traitement linguistique de l'information documentaire*. Paris : Entreprise moderne d'édition, 1982, 186p.

CHRISTIN Anne-Marie (dir.) (2012). *Histoire de l'écriture. De l'idéogramme au multimedia*, Paris : Flammarion, 2012, 413p. (Collection Histoire de l'art).

CLAVIER Viviane (2014). *L'organisation des connaissances au prisme du langage, du texte et du discours. Un parcours en recherche d'information*. Mémoire pour l'obtention de l'habilitation à diriger des recherches en sciences de l'information et la communication, 2 volumes, soutenue le 10 mars 2014 à l'Université Stendhal-Grenoble3, 238p. Disponible sur : <tel:01208271> (consultée le 15 mai 2019).

CLAVIER Viviane et PAGANELLI Céline (2015). Activités informationnelles et organisation des connaissances : résultats et perspectives pour l'information spécialisée. *Les cahiers de la Société Française des Sciences de la Communication*, septembre 2015, n°11, p. 170-175.

CLAVIER Viviane et PAGANELLI Céline (dir.) (2013). *L'information professionnelle*, Paris : Lavoisier Hermès Sciences Publications, 2013, 243p. (Collection systèmes d'information et organisations documentaires).

Conférence permanente des directeurs·trices des unités de recherche en sciences de l'information et de la communication (CPDirSIC) (2018). *Dynamique des recherches en sciences de l'information et de la communication*, septembre 2018, 190p. Disponible sur : <http://cpdirsic.fr/wp-content/uploads/2018/09/dynamiques-des-recherches-sic-web-180919.pdf> (consultée le 15 mai 2019).

GNOLI Claudio (2008). Ten Long-Term Research Questions in Knowledge Organization. *Knowledge Organization*, 2008, vol. 35, n°2/3, p. 137-149.

- GUIRAUD Pierre (1959). *Problèmes et méthodes de la statistique linguistique*, Dordrecht : D. Reidel, 1959, 145p.
- HABERT Benoit, NAZARENKO Adeline et SALEM André (1997). *Les linguistiques de corpus*, Paris : Armand Colin, 1997, 240p.
- HACHOUR Hakim (2015). De la fouille à la visualisation de données : un processus interprétatif. *I2D – Information, données & documents*, 2015, vol. 52, n° 2, p. 42-43.
- HJØRLAND Birger (2015). Theories are knowledge organizing systems (KOS). *Knowledge Organization*, 2015, vol. 42, no. 2, p. 113-128.
- HUDON Michèle et MUSTAFA EL HADI Widad (2010). Organisation des connaissances et des ressources documentaires. De l'organisation hiérarchique centralisée à l'organisation sociale distribuée. *Les Cahiers du numérique*, 2010, vol. 6, n° 3, p. 9-38.
- IBEKWE-SANJUAN Fidelia (2007). *Fouille de textes. Méthodes outils et applications*, Paris : Lavoisier Hermès Sciences Publications, 2007, 352p. (Collection systèmes d'information et organisations documentaires).
- LALLICH-BOIDIN Geneviève et MARET Dominique (2005). *Recherche d'information et traitement de la langue : fondements linguistiques et applications*, Villeurbanne : Presses de l'Enssib, Novembre 2005, 288p. (Les Cahiers de l'Enssib).
- LAMY Jean-Baptiste (2017). *Représentation, iconisation et visualisation des connaissances : Principes et applications à l'aide à la décision médicale*. Habilitation à diriger les recherches en informatique, Normandie Université, Université de Rouen, 2017. Disponible sur : <tel-01849832> (consultée le 15 mai 2019).
- LEBART Ludovic, SALEM André (1994). *Statistique textuelle*, Paris : Dunod, 1994, 342p.
- MOREUX Jean-Philippe (2016). Innovative Approaches of Historical Newspapers: Data Mining, Data Visualization, Semantic Enrichment. Paper presented at: IFLA WLIC 2016 – Columbus, OH – Connections. Collaboration. Community in Session S21 - Satellite Meeting: News Media. In: News, new roles & preservation advocacy: moving libraries into action, 10-12 August 2016, Lexington, KY, USA. Disponible sur : <https://hal-bnf.archives-ouvertes.fr/hal-01389455/document> (consultée le 15 mai 2019).
- PAPY Fabrice (2005). Au-delà de la transfiguration du catalogue. *Bulletin des bibliothèques de France (BBF)*, 2005, n° 4, p. 5-12.
- POIBEAU Thierry (2014). Le traitement automatique des langues pour les sciences sociales. Quelques éléments de réflexion à partir d'expériences récentes. *Réseaux*, 2014, vol. 188, n° 6, p. 25-51.
- POLITY Yolla, HENNERON Gérard et PALERMITI Rosalba (ed.) (2003). *L'organisation des connaissances. Approches conceptuelles*, Actes du 4<sup>e</sup> congrès ISKO-France, 3 et 4 juillet 2003, Paris, Budapest, Turin : L'Harmattan, 2005, 266p. (La librairie des humanités)
- POUDAT Céline, CLEUZIOU Guillaume et CLAVIER Viviane (2006). Catégorisation de textes en domaines et genres : complémentarité des indexations lexicale et morphosyntaxique. *Document numérique*, 2006, vol. 9, n°1, p. 61-76.
- REES D. and LARAMEE R.S (2019). A survey of Information Visualization Books. *Computer Graphics forum*, 2019, vol. 38, Issue 1, p. 610-646.
- REYMOND David (2016). Introduction. Visualisation de données. *Les Cahiers du numérique*, 2016, vol. 12, n° 4, p. 9-18.
- ROUAULT Jacques (1987). *La linguistique automatique. Applications documentaires*. Berne, Francfort/Main, Paris : Peter Lang, 1987, 309p.
- SÉBILLOT Pascale (2002). *Apprentissage sur corpus de relations lexicales sémantiques. La linguistique et l'apprentissage au service d'applications du traitement automatique des langues*.

Mémoire pour l'obtention de l'habilitation à diriger des recherches en informatique, Université de Rennes 1, Décembre 2002, 107p.

SOSIŃSKA-KALATA Barbara (2012). L'efficacité des systèmes d'organisation des connaissances : un point de vue praxiologique. *Études de communication*, 2012, vol. 39, n° 2, p. 155-171.

TUTIN Agnès (1997). Coder les collocations dans un lexique formel pour le TALN. *Revue française de linguistique appliquée*, 1997, vol. ii, n°1, p. 43-58.

VALETTE Mathieu et EGGLE Eensoo (2014). Approche textuelle pour le traitement automatique du discours évaluatif. *Langue française*, 2014, vol. 184, n°4, p. 109-124.

VAN SLYPE Georges (1977). *Conception et gestion des systèmes documentaires*, Paris : Les éditions d'organisation, 1977, 261p. (Collection systèmes d'information et de documentation).

YIKUN Liu et ZHAO Dong (2016). *La datavisualisation au service de l'information*, Paris : Pyramid, 2016, 239p.