Knowledge Organization, Data and Algorithms: The New Era of Visual Representations*

Viviane Clavier

GRESEC—Université Stendhal Grenoble 3—Institut de la Communication et des Médias, 11, avenue du 8 mai 1945—BP 337—38434 Échirolles Cedex,
<viviane.clavier@univ-grenoble-alpes.fr>

Viviane Clavier is Associated Professor and Research Director in information and communication sciences. She is a member of Gresec at the University of Grenoble Alpes and her research focuses on knowledge organization and information seeking in the field of specialized and professional information. These studies focus on the themes of health and nutrition.

Clavier, Viviane. 2019. "Knowledge Organization, Data and Algorithns: The New Era of Visual Representation." *Knowledge Organization* 46(8): 615-621. 35 references. DOI:10.5771/0943-7444-2019-8-615.

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 that contribute to defining new standards and KO representations.

Received: 30 September 2019; Revised: 4 November 2019; Accepted: 4 November 2019

Keywords: knowledge, data, visual representations, images, information

* Presented at ISKO-France 2019: Open Data and Megadata in Social Sciences: New Challenges for the State and Organization of Knowledge?, at Université Paul-Valéry, Montpellier 3, France, October 9-11, 2019.

1.0 Introduction

This contribution is concerned with the place of languages in the process of collecting, representing and organizing knowledge in a context characterised by the massive data production. By suggesting in the title that visual representations are becoming crucial in representing knowledge, we implicitly mean that they have supplanted languages. We are not saying that languages have disappeared, as the use of controlled languages is still relevant to heritage institutions and several fields of specialized information. However, the attention paid to languages has shifted to other knowledge objects (algorithms), other semiotics (data visualization), other levels of representation (textual data), other purposes of knowledge organization (browsing rather than categorization) and other properties (transparency and simplicity rather than precision and exhaustiveness). It is important to notice that the term "language" is not mentioned in the last report published by the IFLA (2018), which is revealing about new trends in information concerning libraries. This paper, therefore, raises a series of questions, puts forward certain hypotheses, looks into the reasons why languages are being called into question, and—if they are being replaced—by what and how. This study has a diachronic rather than historical dimension, which can be seen via research work in information and communication sciences (CPDirSIC 2018), and it also makes reference to our own research in which language, texts and discourse have always played a central role (Clavier 2014). Our presentation is in two parts: the first part shows how visual representations have progressively taken over from language-based representations in knowledge organization systems. The second part shows how this trend observed in research has also spread to civil society and to professionals concerned by information and communication. However, it is clear that languages are still very present, even if they appear to be relegated to a background descriptive function.

2.0 The challenges of knowledge organization: development and permanence

2.1 Applied perspective: languages for knowledge organization systems

The notion of "languages" will be restricted to symbolic systems chosen to encode units of language and meanings, which are used to feed knowledge organization systems (KOS), "tools intended to organize recorded knowledge resources" (Sosińska-Kalata 2012). KOSs include "all kinds of organization schemes ranging from simple alphabetical or slightly structured lists (authority records, glossaries, dictionaries, parts lists, etc.) to hierarchical classification schemes (classification plans, general or specialised classifications, taxonomies, subject headings lists, etc.) or organizations emphasising relations that are not exclusively hierarchical (thesauruses, semantic networks, ontologies, etc.)" (Polity et al. 2005, 13). This perspective places us immediately in what should be referred to as "the narrow sense" of knowledge organization, namely the creation of information products intended to classify, organize and structure knowledge in order to facilitate access to information. This perspective is situated between a "broad" and theoretical conception of knowledge—for an ontological (Gnoli 2008) or social (Hjørland 2015) conception of knowledge—and an "intermediate" approach centered on the applications and forms of knowledge mediation that can be observed in professional situations (Clavier and Paganelli 2013, 2015).

What is common to documentary languages, classifications, folksonomies, ontologies, etc. and today's "data" is that they use processing methods in order to represent content. Thus, at any point in time, knowledge is the result of choices and transformations governing its selection, perhaps its standardization and its categorization. The huge quantity of data to be processed is nothing new—even in the 1970s the development of document processing was based on the idea that automation would help reduce the costs of manually indexing the constantly increasing documentation that professionals could no longer handle (Chaumier 1982). What we intend to question here is the nature of the knowledge selected with regard to computer processing choices.

2.2 From the 1980s to today: the end of languages?

From 1980 to 1990, the development of full text information systems and other applications such as linguistic technology-based knowledge extraction (Chaudiron 2007), enabled the analysis of languages from the standpoint of linguistics (Van Slype 1982). The encounter between computational linguistics and documentation (Rouault 1987) brought together "these two disciplines even though they were separated by a wall" (Van Slype 1982, 87). Viewed from the logico-syntactic perspective of knowledge that could be processed by computer programs, natural languages were the subject of theoretical debates and considerable socioeconomic stakes. The questions raised were as follows (Lallich-Boidin et Maret 2005): 1) What is the relevant unit of meaning: the word or the morpheme?; 2) What advantages and disadvantages do free languages have in comparison with controlled languages?; and, 3)How can meaning be represented accurately while taking into account ambiguities and implications? It is necessary to use external resources in order to define indexing languages (thesauri, controlled vocabularies) or to the documents themselves—a question that raised the issue of faithful indexing in light of cost savings in order to maintain lexical resources. This approach conceived languages as being worked with and by computational linguists, and questioned in relation to language units. It focused on a detailed description of language, based on a compositional approach to meaning led by syntax in the form of logical representations, and was replaced by another approach based on an empirical representation of meaning in which the aim was no longer to understand the language but to describe textual corpora.

The 1990s thus represented a turning point characterised by the web development, paving the way to corpus linguistics (Habert et al. 1997), an approach that shifts the focus towards the creation and description of textual data. And so began the period of "large corpora" intended for annotation by exploration tools as a means of assessing the representativeness of language phenomena. From this standpoint, emphasis was again put on the "vast" character of data and text annotation or "tagging" aimed to collect typologies based on correlations of linguistic features (Biber 1993). These annotation methods, which were often manual or semi-automatic, combined with more robust information processing methods such as learning-based text mining, aimed to perform supervised or unsupervised automatic classification tasks. The knowledge resulting from corpus annotation can be morpho-syntactic categories (Poudat et al. 2006) or collocations (Tutin 1997), with phraseology having benefited from the joint development of symbolic methods and probabilistic methods via automatic learning. Thus, unlike knowledge produced by computational linguistics, which aims to automate applications, the knowledge resulting from annotations produced by corpus linguistics "is based on an iteration between the analysis of computer outputs and 'human' consultation of texts or fragments of texts, for example concordances" (Valette and Egle 2014). The production of knowledge thus forms part of a tool-based approach aimed at exploring data in order to pinpoint linguistic recurrences (idioms, expressions, etc.). This perspective led to the development

of methods based on "opportunistic" linguistic approaches, i.e., those that exploit automatically identifiable surface phenomena.

Over time, the empirical approach became entrenched and enthusiasm for numerical processing methods increased. The development of data analysis tools and the generalisation of methods intended for social sciences and the "general public" was accompanied, as Dominique Boullier points out (Boullier 2015), by "the popular belief that announced the end of science and scientific theories." Thus began a new period in which big data "would be the effective measurements of reality" (*ibid.)*. According to this view, data are based on the collection of bags of words obtained from the web that are characterised by their volume and can be visually represented in two-dimensional spaces. It is then a matter of interpreting and discovering this knowledge, the status of which is not always very clear; does knowledge represent themes, subjects, key words, discourses, lexicometric universes, etc.? The drawback of these tools often lies in the opacity surrounding the data pre-processing stages, the choice of categories (both lexical and textual) or units of discourse. As pointed out by Pascale Sébillot in 2002, mixed methods have indeed been developed, but in limited fields where the aim is no longer to understand texts in detail but to obtain representations of meanings that are useful for precise applications (Sebillot 2002). Thus, rule-based systems are used to develop resources for analysing sentiments and opinions (Poibeau 2014), so that it is possible to annotate data intended to train classifiers. Other applications also rely on the profusion of data, such as machine translation, data journalism and fact checking. However, as pointed out by Thierry Poibeau (ibid.), the part connected with computational linguistics is variable, and "the recurrent difficulties concern system adaptability, the time required to develop resources for a new field and the availability or lack of a sufficient number of examples." The researcher also indicated that other frequent questions concern the definition of the information sought and the quality of the results obtained.

For a number of years now there has been considerable interest in visual representation, as revealed by a recent review of works on the subject carried out by researchers at the University of Swansea (Rees and Laramee 2019). The authors' research goes back as far as 1967, to the *Semiology of Graphics* by Jacques Bertin, one of the first monographs devoted to the question. According to them, the following forty years were not particularly prolific, but since the early 2000s there has been a plethora of publications, to the point where it is difficult to keep track of them (610). Visual representation must be understood as both a process of creating visibility and as the result of that process. In the former case, it is "the result of the instrumented exploration of masses of data that become suitable for generating indicators, maps, etc." (Reymond 2016, 11). In the latter, the produced visual representations form the material for computer graphics, laid out in such a way as to form a language in its own right. However, what is visualized may be data, information or knowledge. "Data visualization" is the generic term used to designate the tools of "dataviz," involving "a semiotic transformation between the results of a data analysis (numerical, categorial, textual) and a graphic representation" (Hachour 2015). The "visual representation of information" is a branch of computing that "uses visual and interactive representations of abstract data on a computer to amplify cognition" (Lamy 2017, 76 ff) and has six possible goals: displaying a large quantity of data, facilitating the search for given information, detecting patterns, enabling visual inferences, monitoring the occurrence of events and exploring data sets. The graphic representations obtained in this way serve several applications intended to visualize information, such as journalistic or documentary information. Thus, data visualization for information (Yikun et Zhao 2016) is a means of discovering journalistic news by establishing "new interactions" and presenting them visually. As far as documentary information is concerned, it may be visualized by maps that help in browsing classifications (Dewey), directories (Rameau) or catalogues (Papy 2005). Lastly the "visual representation of knowledge" falls within the field of research into artificial intelligence and is based on the representation, modelling and visualisation. According to the computer science researcher Jean-Baptiste Lamy (Lamy 2017, 12), a distinction must be made between the iconic visual representation of knowledge, which involves translating knowledge by using an iconic language including a pictogram glossary and a grammar, and a structural visual representation of knowledge, which involves representing the structure of knowledge graphically using the techniques of visual representation.

3.0 Knowledge dissemination: priority of visual representations for many actors

The choice of verbal or visual semiotics to represent knowledge in information systems is not simply a matter for science but falls within a much wider spatio-temporel, socio-economic and cultural context. Visual representations are now the predominant modes of expression in our society, as evidenced by computer graphics, fixed and animated images (videos) in the field of public and scientific communication, in the media and on the web. To remove any ambiguity, it is not a question of demonstrating that we live in "a world of images" as suggested by the philosopher Franck Robert in his introduction to the joint publication "Philosophies de l'image," but rather that the importance assumed by visual representations—or "visuals"

as they are referred to by communication specialists—as forms of semiotization influences the ways in which knowledge is represented and organized. The rest of this article will present a few trends observed among actors who, for one reason or another, play an important role in the dissemination of knowledge representation standards or formats. These examples—which are not derived from a reasoned corpus—illustrate the fact that knowledge is always rooted in particular periods and places and probably depends on fashions.

3.1 Promotion of images in web publication tools and techniques

Images are promoted by the web professionals who define the writing standards and techniques aimed at optimising the visibility of websites via their referencing and positioning in search engines. There are very many sites aimed at web publishers giving recommendations on how to write for the web and setting out rules on concision, simplicity, content structure, addresses, etc.[1] With regard to images, they may be fixed or animated, such as videos. On the website Annei, images and text, "always considered to be rivals," are now presented as complementary, subject to a few adjustments intended to prevent images being restricted to a secondary role. The site advises combining "alternative text," a "concise, descriptive textual equivalent" along with images that do not exceed "250 characters," or "ensuring that video soundtracks will be made accessible by a retranscription in text format that can be detected by search engines." As far as the choice of images is concerned, it also gives recommendations on their content: "Choose images that have an informative value." This advice is aimed at eliminating images that are purely illustrative. Images considered to be informative include: "computer graphics, diagrams, photographs taken in real situations, providing more than a visual taken from a data bank and is of purely decorative value" (*ibid.*). In this last instance, it advises adding "a concise legend that gives meaning to the image, whenever possible," in order to reinforce "the image's impact rating."

The advice given with regard to publishing on the social media is that images should even replace text on the grounds that an image simply requires a click to score a hit, thus guaranteeing traffic that generates profits or viewership. This recommendation appears to be connected with users' observed habits: they click on images to browse the internet:

> Pinterest and Facebook users are accustomed to zigzag browsing. One click leads to another click, an image leads to a link where there are other images, which also lead to other links. Images are like Tom Thumb's little pebbles. If they are scattered intelligently, they will attract far more traffic to a website than a simple 'quote.'
>
> When you realise that photos posted on Facebook can generate 53% more 'likes' on a post, you make the most of them. Users are far more inclined to like or follow a brand if it is active and shares images. Textual promotion is a thing of the past.

Web writing standards for the social media now advise choosing visual semiotics: it is better to "show" than to "tell," words are "there to underscore the image," "visual stratagems are to be used to win visitors," etc. Even so, the eternal question of document description, representation and indexing is still raised in the context of information searches in the form of user requests. Ultimately, then, images always lead back to the question of languages and the choice of language to describe them. This question is dealt with more specifically on sites intended to enhance the visibility of websites in search engines (search engine optimization) or on the social media (social media optimization) or specifically in engines dedicated to marketing (search engine marketing). Reference is made to W3C, which defines web languages and standards and gives the image formats accepted by search engines, description tags and advice on describing the alt and title attributes for presenting "the content of an image clearly and concisely."

> The alternative text of images: Google also references images! Always remember to complete the alternative text (alt tag) of the images in your articles. Introduce your main keyword within a short expression. The alt tag regularly omitted is that of the image on the first page, which weakens image referencing. To avoid this, fill in the Alternative text field when you define your first page image.

It is clear that images, far from replacing text, once again raise the question of textual representation, the choice of words and their descriptive virtue, reversing the roles of "illustrative" images in favour of text. From the point of view of information searching, these trends have led to the development of research into the automatic classification of images, facial recognition, and the production of languages for describing and annotating images, such as ontologies. Indeed the "Google Image" search engine is becoming increasingly popular with users to the extent that natural image referencing now appears to be a priority.[2]

3.2 Development of data visualization tools for the general public

Hitherto reserved for specialists in multidimensional descriptive statistics, certain methods such as principal component analysis (PCA) and correspondence factor analysis (CFA) have been widely used in the humanities and social sciences to study literary or political texts from different viewpoints: lexicon, vocabulary, style, themes, etc. An extensive literature devoted to textual statistics has been built up since the late 1950s, the common theme of which is the production of knowledge organization systems in the form of verbal semiotics: dictionaries (Guiraud 1959), indexes, concordances, repeated segments, textual time series, parallel corpora, etc. (Salem et Lebart 1996). In these works, visual representations—entire lexical tables, figures representing factorial planes, etc.—are not a goal in themselves but are considered to be an "aid in reading a series of texts" (135 ff). These introductory works give guidelines on: "How to interpret distances," "Examples of applications," "Reading a figure," etc. It is thus clear that the interpretation of textual data is subject first of all to lexicometric reorganization in the form of visual representations, which are the result of a set of processing operations and choices concerning the size of the textual corpora, their representativeness and segmentation, the definition of counting units, etc. As a follow-up to these methods, text mining includes statistical data processing and tasks such as supervised or unsupervised classification (Ibekwe-SanJuan 2007). Developed by computer scientists specialising in learning, these methods approach texts as data intended to train classifiers, whose scores are measured and compared. Applied research as information retrieval, pattern detection, scientific monitoring, etc. are always envisaged, so that the text in the strict sense and considered in its production conditions, is not the focus of attention. It is this transition from viewing text as sacred to viewing it as an instrument, a training corpus, that appears to be at the origin of an interpretation centred on understanding trends, regularities, mass effects, etc. and less on what is of the order of the invisible, which can only be retrieved through knowledge of the texts.

At the same time, dataviz tools are becoming far more commonplace as many actors require analyses of discourse on the social media, blogs, Twitter, forums, etc. Thus, for economic reasons, survey institutions have introduced so-called passive methods, inspired from big data, which are qualitative, can be automated and used to complement quantitative suveys. There is a wide social demand: marketing, political polls, consumption, etc. There is a multitude of websites that list dataviz tools, distinguishing between various families: mental maps, visual representation of relational data networks, information mapping, etc. Their success would appear to be due to their transparency, simplicity and faithfulness to the data, as can be seen in this example of an announcement presenting the advantages of dataviz for representing textual or numerical data.

> Dataviz, or data visualization, is a practice that we encounter on a daily basis without realising it, just by opening a newspaper or watching the television. The simplest example is the survey. With the digital age, it has become a powerful communication tool.
>
> Dataviz: a definition that is easy to understand.
>
> In a society that is increasingly attracted to graphics, data visualization takes precedence over raw data. It helps to throw light on information that is apparently complex or is submerged in a large quantity of parameters. The term dataviz thus refers to a set of visual representations of these raw data. What is its principal objective? To throw light on a phenomenon by giving it a more ergonomic appearance than a spreadsheet filled with figures.

Transparency refers to the immediately perceptible and understandable character of a message when it is represented in a visual form, which would give it undeniable communicational virtues. Simplicity is the ability of the tools to provide a summary representation in visual form of information that is expressed in a verbal manner, which is thus felt to be more complex. Lastly, faithfulness is the ability of the tools to represent data such as that emanating from individuals without any form of mediation. These properties, which are presented as technical assets, are the subject of ongoing theoretical debates on the differences in semiotic status of pictorial and written signs (Christin 2012). These arguments hark back to the sterile debates on the more transparent character of pictorial signs in comparison with written signs on the grounds that there is no need to use an arbitrary code to interpret an image, in contrast to a text; likewise simplicity no doubt refers to the small number of signs used, as a small set of visual forms is always more concise than an infinite number of verbal signs combined to describe a language. These debates are far from over, and they question the pertinence of knowledge encoding processes without taking into account reception conditions. Lastly, faithfulness is also a question that is widely discussed in the information-documentation field with regard to the merits of free and controlled indexing languages.

The representation of knowledge in visual form is also of interest to communication specialists who approach information design from the aesthetic standpoint. They are required to posses new dataviz skills at the crossroads between computing, statistics and communication (Arruabar-

rena 2017). Lastly, the visual representation of knowledge has been lifted to the status of an artistic object worthy of being displayed. For example, the temporary exhibition at the Mundaneum in Mons entitled "Mapping Knowledge" presents the work of Paul Otlet, who is more famous for having created the universal decimal classification system than for his texts and drawings aimed at "summarising what is known" (see Figure 1).

4.0 Conclusion

Focusing on the place of visual representations as knowledge organization systems is part of a long line of research into the choice of vocabularies and how they are organized and structured in order to access knowledge. This contribution illustrates the fact that visual representations obtained through data processing are neither "natural," nor "simple," nor are they "objective" even if they are made by automatic tools and are applied to already existing data (Bullich and Clavier 2018). This observation was made in particular by Michèle Hudon and Widad Mustapha El Hadi with respect to classifications: "There is no natural organization structure and any classification or structure can only be invented and built, never discovered. Furthermore, the establishment of a classification framework or structure is always evidence of an encounter or clash between philosophical, technological, social, economic and political imperatives (Svenonius 1991)" (Hudon and Mustafa El Hadi 2010, 12).

Supported by specialist disciplines in big data processing and encouraged by research and education policies, data visualization is more than just a tool for the knowledge management profession, it is a genuine aid to discovering knowledge. It is no longer a matter of naming knowledge in order to understand it, but rather of showing it in order to make it visible, a perspective that leaves considerable leeway for new actors, new specialisations and new data collection structures that are helping to redefine standards, forms of writing and methods for representing knowledge.

Notes

1. The examples of sites mentioned here are among the top results obtained by Googling "web image and text publication."
2. https://www.1ere-position.fr/blog/optimiser-le-referencement-dimage-sur-google-images/#Google_Images_en_forte_croissance (consulted on 2 May 2019).

References

Arruabarrena, Béa. 2017. "L'expert en dataviz, un métier en transition." *I2D – Information, données & documents* 54, no. 3: 7-8.

Biber, Douglas. 1993. "Using register-diversified corpora for general language studies." *Computational Linguistics* 19: 243-58.

Figure 1. Screen capture of the exhibition devoted to the visual representations of Paul Otlet, obtained from GoogleArts& Culture.

Boullier, Dominique. 2015. "Les sciences sociales face aux traces du *big data*. Société, opinion ou vibrations ?" *Revue française de science politique* 65: 805-28.

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.*

Chaudiron, Stéphane. 2007. "Technologies linguistiques et modes de représentation de l'information textuelle." *Documentaliste-Sciences de l'Information* 44: 30-39.

Chaumier, Jacques. 1982. *Analyse et langages documentaires. Le traitement linguistique de l'information documentaire*. Paris: Entreprise moderne d'édition.

Christin, Anne-Marie. 2012. *Histoire de l'écriture. De l'idéogramme au multimedia.* Paris: Flammarion.

Clavier, Viviane. 2014. *L'organisation des connaissances au prisme du langage, du texte et du discours. Un parcours en recherche d'information*. Habilitation à diriger des recherches en sciences de l'information et la communication, Université Stendhal-Grenoble3.

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*, 11: 170-75.

Clavier, Viviane et Paganelli, Céline. 2013. *L'information professionnelle.* Paris: Lavoisier Hermès Sciences Publications.

CPDirSIC (Conférence Permanente des Directeurs·trices des unités de recherche en Sciences de l'Information et de la Communication). 2018. *Dynamique des recherches en sciences de l'information et de la communication*.

Gnoli, Claudio. 2008. "Ten Long-Term Research Questions in Knowledge Organization." *Knowledge Organization* 35: 137-49.

Guiraud, Pierre. 1959. *Problèmes et méthodes de la statistique linguistique.* Dordrecht: D. Reidel.

Habert, Benoit, Nazarenko, Adeline et Salem, André. 1997. *Les linguistiques de corpus.* Paris: Armand Colin.

Hachour, Hakim. 2015. "De la fouille à la visualisation de données: un processus interprétatif." *I2D – Information, données & documents* 52, no. 2: 42-43.

Hjørland, Birger. 2015. "Theories are knowledge organizing systems (KOS)." *Knowledge Organization* 42: 113-28.

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* 6, no. 3: 9-38.

Ibekwe-Sanjuan, Fidelia. 2007. *Fouille de textes. Méthodes outils et applications*, Paris: Lavoisier Hermès Sciences Publications.

IFLA. 2018. *Riding the waves or caught in the tide? Navigating the evolving information environment. Insights from the trend report.*

Lallich-Boidin, Geneviève and Maret, Dominique. 2005. *Recherche d'information et traitement de la langue: fondements linguistiques et applications,* Villeurbanne: Presses 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 des recherches en informatique, Université de Rouen.

Lebart, Ludovic et Salem, André.1994. *Statistique textuelle.* Paris: Dunod.

Papy, Fabrice. 2005. "Au-delà de la transfiguration du catalogue." *Bulletin des bibliothèques de France* 4: 5-12.

Poibeau, Thierry. 2014. "Le traitement automatique des langues pour les sciences sociales. Quelques éléments de réflexion à partir d'expériences récentes." *Réseaux* 188, no. 6: 25-51.

Polity, Yolla, Henneron, Gérard et Palermiti, Rosalba. (ed.) 2003. *L'organisation des connaissances. Approches conceptuelles*. Actes du 4e congrès ISKO-France, Paris: L'Harmattan.

Poudat, Céline, Cleuziou, Guillaume et Clavier, Viviane. 2006. "Catégorisation de textes en domaines et genres: complémentarité des indexations lexicale et morpho-syntaxique." *Document numérique* 9: 61-76.

Rees, Dylan and Laramee, Roberts S. 2019. "A survey of Information Visualization Books." *Computer Graphics forum* 38: 610-46.

Reymond, David. 2016. "Introduction. Visualisation de données." *Les Cahiers du numérique* 12, 4: 9-18.

Rouault, Jacques. 1987. *La linguistique automatique. Applications documentaires.* Berne: Peter Lang.

Sébillot, Pascale. 2002. *Apprentissage sur corpus de relations lexicales sémantiques. La linguistique et l'apprentissage au service d'applications du traitement automatique des langues.* Habilitation à diriger des recherches en informatique, Université de Rennes 1.

Sosińska-Kalata, Barbara. 2012. "L'efficacité des systèmes d'organisation des connaissances: un point de vue praxiologique." *Études de communication* 39: 155-71.

Tutin, Agnès. 1997. "Coder les collocations dans un lexique formel pour le TALN." *Revue française de linguistique appliquée*, vol. ii: 43-58.

Valette, Mathieu et Egle, Eensoo. 2014. "Approche textuelle pour le traitement automatique du discours évaluatif." *Langue française* 184: 109-24.

Van Slype, Georges. 1977. *Conception et gestion des systèmes documentaires.* Paris: Les éditions d'organisation.

Yikun, Liu et Zhao, Dong. 2016. *La datavisualisation au service de l'information*. Paris: Pyramid.