

Rapport n°17-37
JUN 2021

LA VISUALISATION INTERACTIVE DE LA JURISPRUDENCE DE LA COUR DE CASSATION (VICO)

Outil : <http://visualex.org/>

Note de synthèse

Sous la direction de : Audilio Gonzalez-Aguilar, maître de conférences, Institut Méditerranéen des Sciences de l'Information et de la Communication (IMSIC), Université de Toulon, Aix-Marseille et Laboratoire Paragraphe EA 349 Université Paris 8.

Ont également contribué à ce rapport de recherche :

- Marc Clément, magistrat, président du Tribunal administratif de Lyon,
- Raphaël Volt, informaticien, docteur en Recherche Numérique à l'Université de Nottingham (Royaume-Uni)
- Stefan du Château, maître de conférences à l'Université de Toulouse 2.

Le présent document constitue le rapport scientifique d'une mission réalisée avec le soutien du GIP Mission de recherche Droit et Justice (convention 17.37). Son contenu n'engage que la responsabilité de ses auteurs. Toute reproduction, même partielle est subordonnée à l'accord de la Mission.

Note de synthèse

LA VISUALISATION INTERACTIVE DE LA JURISPRUDENCE DE LA COUR DE CASSATION (VICO)

Outil : <http://visualex.org/>

I. LES FONDEMENTS ET CONCEPTS DE LA VISUALISATION	5
II. LA VISUALISATION DES DONNEES, DE L'INFORMATION ET DES CONNAISSANCES	9
III LA VISUALISATION DES DONNEES DANS LE DROIT	10
IV LES FONDEMENTS THEORIQUES DE LA RECHERCHE	12
V. LES FONDEMENTS METHODOLOGIQUES DE LA RECHERCHE	14
VI. RESULTATS ET LIVRABLES DE VICO	15
CONCLUSION	19

Introduction

Le projet de Visualisation Interactive de la jurisprudence de la COur de Cassation (VICO) cherche à faire une représentation visuelle interactive et analytique de la jurisprudence présente sur le site de la Cour de Cassation (<https://www.data.gouv.fr/fr/datasets/cass/>). Il répond au défi de la croissance exponentielle de la production des documents juridiques, pour lesquels seront mis en œuvre de nouveaux processus, modèles et techniques de représentation et de traitement. Il donne l'opportunité, à partir de l'Open data, de produire des artefacts réutilisables, c'est-à-dire un dispositif de jurisprudence. Avec VICO, les praticiens du droit disposent de l'information jurisprudentielle exhaustive pour bâtir des stratégies juridiques pertinentes nécessaires à la gestion des litiges, à l'appréhension des jugements ou aux travaux de recherche.

Le projet a comme point de départ la notion fonctionnelle du document. Dans cette approche, le document a d'une part, une fonction de communication, et d'autre part, une fonction de mémoire :

- La jurisprudence est une trace du passé et une composante de la mémoire, qui désigne à la fois d'une part, les connaissances juridiques concernant un individu particulier dans les litiges, et d'autre part les catégories juridiques traitées dans chaque arrêt, ce qui relève d'un collectif, d'une institution : le savoir juridique assuré par la jurisprudence de la Cour de Cassation.
- La jurisprudence a le pouvoir de renvoyer au passé grâce aux relations qu'elle entretient avec le contexte juridique donné par le réseau sémantique des métadonnées qui constitue la base de la visualisation de la jurisprudence. Ce sont les relations du réseau jurisprudentiel qui changent avec le temps et non la jurisprudence. Ce qui change c'est la façon que nous avons de l'appréhender¹.
- A l'ère des humanités numériques, la visualisation des données jurisprudentielles est progressivement conçue comme essentiellement procédurale, à décrypter sous la forme de réseaux sémantiques. En effet, Saussure et Derrida pensaient que le texte contenait même des images mentales de mots (Ratnapala, 2013 p. 25). Il existe une forme de connaissance qui peut être découverte à partir des données visuelles : le tissage des métadonnées sémantiques contenues dans le document numérique. Une telle organisation génère un « infomapping » de schémas de réseaux sémantiques des métadonnées contenues dans le texte jurisprudentiel.

Nous avons choisi, dans une première partie, de réaliser une approche historique de la visualisation appliquée du droit depuis le Moyen-Âge pour montrer son importance dans la construction de la connaissance, puis d'en donner les différentes définitions ; d'aborder, dans une seconde partie, l'historique, le processus et les aspects taxonomiques de la visualisation des données, de l'information et de la connaissance, puis de contextualiser le domaine de la visualisation dans la discipline du droit dans la troisième partie. Les parties suivantes, quatre et cinq, présentent le positionnement du projet par rapport à l'état de l'art et la méthodologie utilisée pour atteindre les objectifs qui sont les nôtres. Enfin, la dernière partie livre les phases de la réalisation et le résultat de la visualisation de la base de données jurisprudentielles CASS.

¹ Cet exemple : « Un film de science-fiction des années trente vu aujourd'hui nous montre le passé parce qu'il permet de nous faire voir que le contexte qui l'entoure et les modalités d'appréhension de ce genre fictionnel ne sont plus les mêmes. » (Jussi, 2013 p. 10)

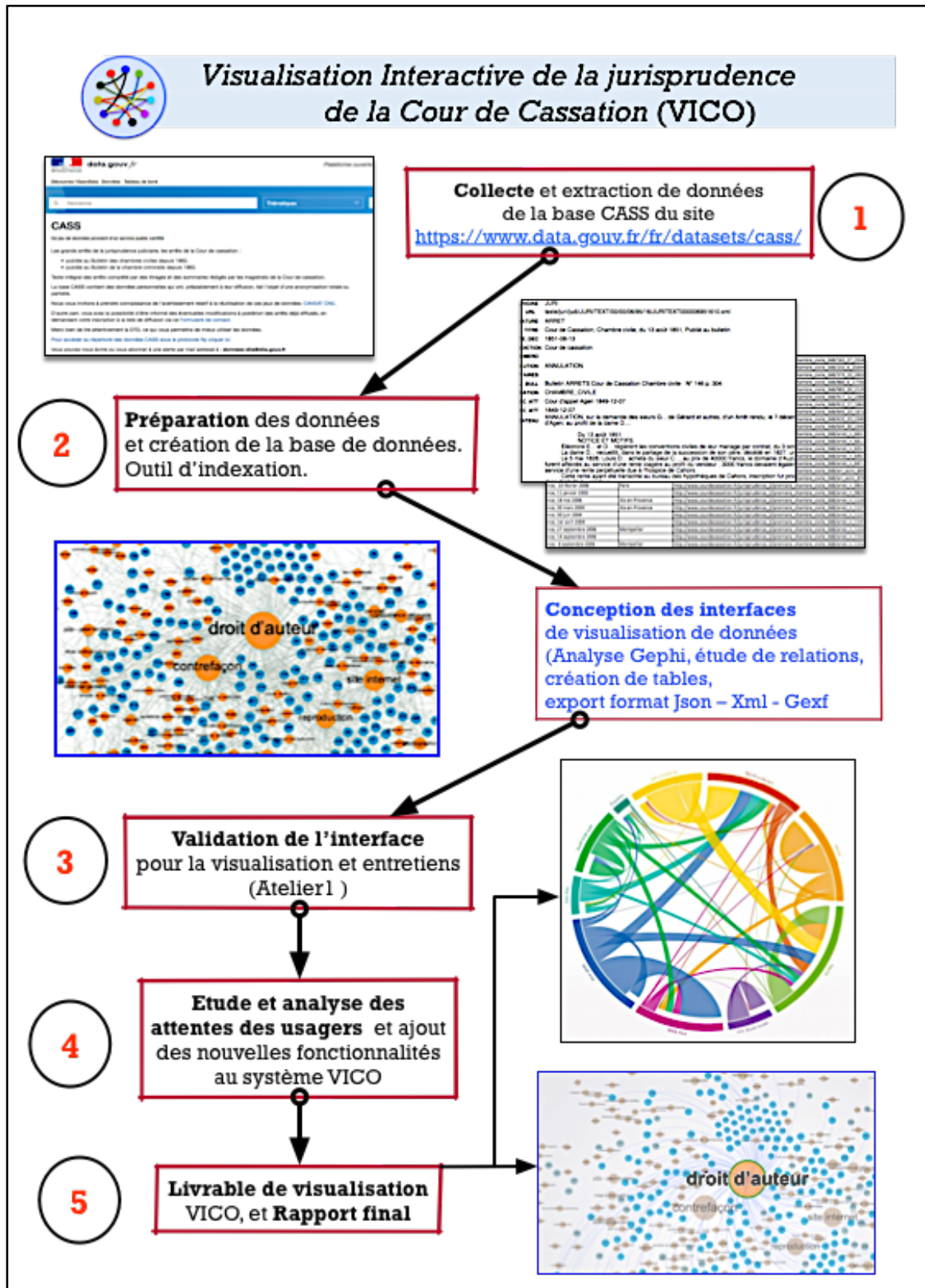


Figure 1: Projet VICO – Visualisation de la jurisprudence Cour de Cassation. (Gonzalez, 2019)

I. Les fondements et concepts de la visualisation

Histoire de la visualisation

Nous avons choisi de nommer ce projet VICO en hommage au philosophe et juriste italien Giambattista Vico² considéré comme le précurseur de la pensée complexe et de l'épistémologie constructiviste, qui correspondent aux hypothèses et méthodologie de recherche de notre projet. La visualisation entendue comme « la représentation graphique de l'information quantitative » a des racines profondes. Ces racines s'étendent aux histoires de la cartographie thématique, des graphiques statistiques et de la visualisation des données, qui sont liées les unes aux autres. Les racines des représentations mentales remontent à la préhistoire, en témoignent les premières images rupestres. Dès l'Antiquité, les Grecs élaborent des méthodes de mnémotechnique et Aristote conclut que « *l'âme ne pense jamais sans une image mentale* »³. Au Moyen-Âge, l'image comme « substitut du langage » se retrouve inscrite dans les vitraux et dans les manuscrits et constitue sous diverses formes un outil de mémorisation et de visualisation des connaissances.

Elle est également liée à l'essor de la pensée statistique tout au long du XIX^e siècle et à l'évolution de la technologie au XX^e siècle. « *Depuis la surface, on peut voir le fruit actuel ; il faut regarder en dessous pour voir son origine et sa germination. Il y a certainement eu beaucoup de choses nouvelles dans le monde de la visualisation ; mais à moins de connaître son histoire, tout peut sembler nouveau.* » (Friendly, 2001). Depuis les systèmes d'information juridique médiévaux, la visualisation prend sa place en tant qu'outil de mémoire et transmission de la connaissance juridique. Elle garantit l'acquisition de la complexité juridique. Dans notre travail, nous analysons plusieurs exemples de visualisation au Moyen Âge européen et l'importance des images dans le dispositif de la mémoire du droit médiéval⁴, puis comment une véritable discipline émerge à partir du XIX^e siècle.

Une des premières représentations juridiques est la métaphore de l'arbre, tels les arbres de consanguinité et d'affinité (*arbores consanguinitatis* et *arbores affinitatis*⁵) qui font leur apparition dès le IX^e siècle. Cette représentation a pour objet de clarifier, de rendre visuellement explicites les degrés de parenté qui unissent les membres d'une même famille, en droit romain et en droit canonique. Avant de se fixer à partir du XVI^e siècle, la métaphore de l'arbre se nourrit à de multiples sources : religieuse, arbre du bien et du mal ; philosophique, arbre de la connaissance, juridiques, arbres représentant les règles de mariage ou de succession⁶. Le *miroir des saxons* (*Sachsenspiegel*), produit entre 1295 et 1371 nous présente la circulation des connaissances juridiques à travers le temps : le « *texte et l'image loin d'être isolée, la visualisation et l'écriture interagissent et interfèrent constamment, dans une réinvention perpétuelle de leurs relations, leur centre de gravité oscillant d'un point à l'autre dans le champ de la communication* »⁷.

Le manuscrit le *Vidal Mayor* nous montre la circulation des métiers dans la création de documents juridiques et constitue un cas unique de manuscrits illustrés en Espagne dans lequel un thésaurus visuel du droit est réalisé pour la première fois. Ce travail matérialise entre autres, la circulation des métiers à l'époque médiévale et conserve la mémoire du droit espagnol. *L'arbor actionum* établit un schéma structuré de connaissances des actions justiniennes et nous permet de

² Vico G. (2001 [1744]), *L'Esprit de la Cité*. Traduit et présenté par Alain Pons, Paris : Fayard.

³ Aristote, De l'âme • 432 a, 17 : «... l'âme ne pense jamais sans une image mentale » ;

• 431 b, 2 : «... la faculté pensante pense ses formes en images mentales » ;

• 432 a, 9 : « Personne ne pourrait jamais apprendre ou comprendre quoi que ce soit sans la faculté perceptive ; même quand on pense spéculativement, on doit avoir une image mentale avec laquelle penser. »

⁴ Voir Robert Jacob, *Images de la Justice. Essai sur l'iconographie judiciaire du Moyen Âge à l'âge classique*, Paris, Le Léopard d'Or, 1994, 256 p., 18.

⁵ Hermann SCHADT, *Die Darstellungen der 'Arbores Consanguinitatis' und 'Arbores Affinitatis', Bildschemata in juristischen Handschriften*, Tübingen, E. Wasmuth, 1982, p. 224, 229-233, 241, n. 48, 264, pl. 93, 94

⁶ Voir <http://classes.bnf.fr/arbre/analyse/index.htm>.

⁷ Wetzell, René, & Flückiger, Fabrice, *Au-delà de l'illustration, Texte et image au Moyen Âge, approches méthodologiques et pratiques*, Zurich : Chronos, 2009 (Medienwandel - Medienwechsel - Medienwissen, 6) pags. 9 y siguientes

comprendre la circulation des connaissances juridiques en Europe entre le V^e et XIII^e siècles. « *L'arbor actionum* » fonde un nouveau paradigme de visualisation, puisqu'il s'agit d'un modèle de pensée visuelle où la représentation du savoir juridique prend toute sa dimension comme dispositif de transmission de la connaissance du droit par la représentation visuelle et comme moyen de preuve dans la procédure judiciaire. L'image, le schéma, constituent un élément fondamental dans la reconnaissance de droits dans le procès. La visualisation de l'information juridique a ses racines au Moyen Âge. En fait, les manuscrits juridiques du XI^e siècle et des siècles suivants contiennent de nombreux diagrammes ou images qui représentent visuellement les données juridiques. Enfin, une application précise de la visualisation en droit est son application dans l'enseignement du droit. Dans deux ouvrages de Johannes Buno (1617-1697), le *Memoriale juris civilis*⁸ romani paru en 1673 et le *Memoriale codicis justiniani*⁹ paru en 1674 les schémas et dessins cherchent à aider les étudiants et praticiens à mieux retenir le contenu complexe du *Corpus juris civilis* qui contient des *tabulae* illustrées.

A partir de XII^e siècle, les juristes établissent une hiérarchie entre les modes de preuves. Au sommet de cette hiérarchie on trouve la preuve notoire (notorium). La preuve par la vue est reconnue comme la meilleure des preuves en même temps qu'elle disparaît de la réflexion, à cause de son caractère même d'évidence. Les schémas géométriques de Bartole de Sassoferrato (1355) dans le traité du droit la Tibériade (Tiberiadis¹⁰) sont aussi des modèles théoriques pour résoudre des problèmes pratiques que peuvent rencontrer le juge ou le notaire. A partir du XVI^e siècle, l'utilisation de l'image en droit connaît un véritable déclin dû à l'apparition de l'imprimerie qui rend plus difficile et coûteuse l'inclusion d'images dans les livres. « Le droit à ce moment-là se séparait de l'image, il devenait une science grise » (Jacob, R. 1994 p194). La visualisation de données en tant que discipline s'est développée dans le domaine des statistiques dès le XVIII^e siècle grâce aux travaux sur les représentations graphiques de l'écossais William Playfair appuyé par des mathématiciens Bayes et Laplace.

Jusqu'au XIX^e siècle, la Data Visualisation s'est construit un vocabulaire mais il lui manquait encore des règles communes – c'est-à-dire une grammaire. En 1920 - Otto Neurath, philosophe, sociologue et économiste autrichien, développe un langage visuel à base de pictogrammes : l'Isotype (International System of Typographic Picture Education). En 1967, paraît la première édition du traité fondateur de Jacques Bertin : « *Sémiologie graphique* ». *Les diagrammes, les réseaux, les cartes* » (Bertin, 1977). Avec cet ouvrage semble s'ouvrir un nouveau champ de connaissances, la « sémiologie graphique », ou science de la représentation graphique des données. Différentes métaphores visuelles théorisées sont présentées, elles permettent des représentations de données dans un espace défini et traduisent simplement les relations complexes qui peuvent lier les données. A ces propos Bertin souligne : « *La Graphique est un moyen de communiquer avec les autres. C'est son emploi le plus connu. Elle sert aussi à poser et à résoudre un problème* ».

⁸ Voir <https://ia800102.us.archive.org/1/items/memorialejurisci00buno/memorialejurisci00buno.pdf>

⁹ Voir https://archive.org/download/memorialecodicis00buno_1/memorialecodicis00buno_1.pdf

¹⁰ *Tiberiadis* est le titre original du traité. Le terme, que Bartole a inventé pour désigner la région du Tibre, depuis sa source jusqu'à son embouchure.

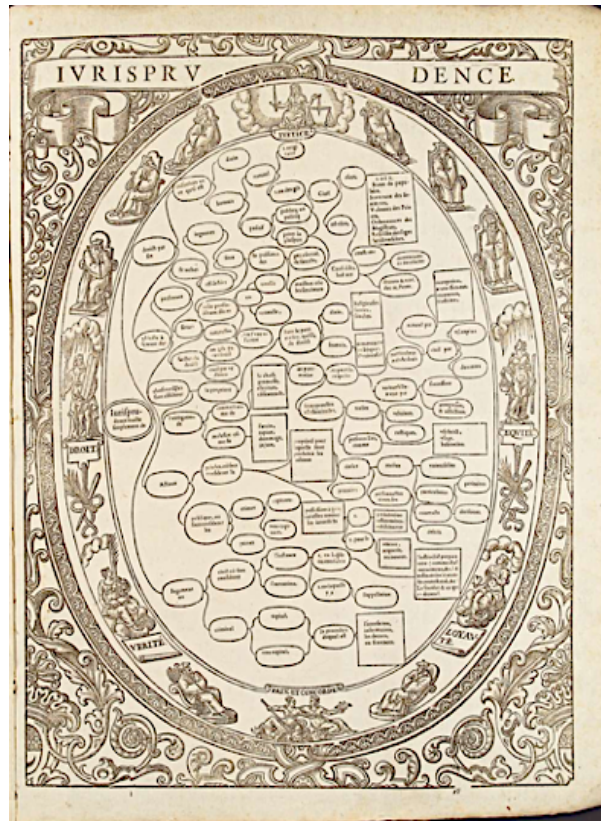


Figure 2 : Christophe de Savigny, *Tableaux accomplis de tous les arts libéraux*, Gourmond, Paris, 1587. License CC 1.0 - Domaine Public

L'image en droit reprend sa place dans le monde juridique à la fin du XX^e siècle avec l'apparition de l'informatique qui permet la création de bases de données et de documents multimédia. La présence du document numérique dans le domaine du droit donne une nouvelle approche du texte juridique, à savoir le fait de considérer le texte juridique comme un ensemble de données. A partir de cet ensemble de données juridiques il devient possible de réaliser des traitements qualitatifs et quantitatifs de l'information juridique et d'en faire sa représentation visuelle. Un long processus de maturation au cours des XX^e et XXI^e siècles lui donneront un nouvel essor, grâce à l'apparition puis la généralisation des traitements informatiques.

Le numérique a permis de récupérer la mémoire du passé mais également les connaissances. L'accroissement exponentiel des documents juridiques (jurisprudence législation, doctrine etc.) a conduit à se trouver en présence d'un ensemble volumineux de données juridiques qui doivent être traitées et analysées. Ainsi nous nous trouvons face au Big Data de l'information juridique qui conduit à une démarche particulière consistant à extraire l'information pertinente de l'ensemble de données. Une des solutions pour traiter le Big data est de transformer un ensemble d'informations complexes en graphiques interactifs, simples à comprendre et à utiliser. La gestion et la capacité du système juridique à faire preuve de résilience¹¹ passe par la mise en place de dispositifs visuels de gestion de la crise générée par le Big Data.

¹¹ Craig (2009, p. 23) suggère que "les objectifs réglementaires et les mécanismes juridiques pour les atteindre devront être centrés sur le concept même du changement". Les juristes ont eu tendance à se concentrer sur la juxtaposition potentielle entre le droit et la résilience en tant que défi complexe de trouver un équilibre entre la recherche de stabilité et de prévisibilité dans les cadres juridiques, la complexité des systèmes socio-écologiques et la flexibilité requise dans la prise de décisions fondées sur la science (Barnes, 2013 ; Cosens, 2013 ; Craig, 2009 ; Ebbesson, 2010 ; Ruhl, 2009). Craig (2009) et Cosens (2013) proposent que des concepts tels que "flexibilité fondée sur des principes" (Craig, 2009) ou "stabilité mesurée" (Cosens, 2013) pourraient être utiles pour aborder l'interrelation complexe entre le cadre juridique (défini par des textes normatifs et des règles fixes et prévisibles) et les écosystèmes dynamiques. Ils proposent que ces concepts pourraient permettre au droit de s'adapter à la transformation continue en permettant l'intégration mesurée de la science pour une résolution plus efficace des conflits. » (Clarvis et al., 2014 p. 4).

Les définitions et caractéristiques de la visualisation

Le mot visualisation a été emprunté à l'anglais "*visualization*"¹². La visualisation des données est une représentation graphique qui exprime l'importance des données. Elle révèle des informations et des modèles qui ne sont pas immédiatement visibles dans les données brutes. C'est un art à travers lequel les informations, les nombres et les mesures peuvent être rendus plus compréhensibles. Il existe autant de définitions de la visualisation des données qu'il y a de définisseurs (nous en avons fait un état des lieux), mais à la racine du terme qui existe depuis plusieurs années, le but est que les données soient visualisées de manière à permettre la compréhension. Nous définissons la visualisation de l'information, plus généralement, comme la communication des données abstraites et pertinentes en termes d'action, grâce à l'utilisation d'interfaces interactives.

« *La visualisation est l'utilisation de représentations visuelles interactives et informatisées de données abstraites pour amplifier la cognition* » (Card, Mackinlay & Shneiderman B. 1999).

Le principal objectif de la visualisation des données est de communiquer des informations de manière claire et efficace par des moyens graphiques (Friedman, 2008). La visualisation des données est un outil si puissant¹³ (Schermann, 2019), car il est : 1. Intuitif, 2. Rapide, 3. Flexible, 4. Perspicace. Il existe trois principaux objectifs de la visualisation, à savoir a) la présentation, b) l'analyse de confirmation et c) l'analyse exploratoire. Voici les trois critères que toute visualisation doit remplir pour être considérée comme une visualisation pragmatique (Kosara, 2007) : 1. Elle doit être réalisée à partir et sur des données non-visuelles. 2. Elle doit produire une image. 3. Le résultat visuel doit être lisible et reconnaissable.

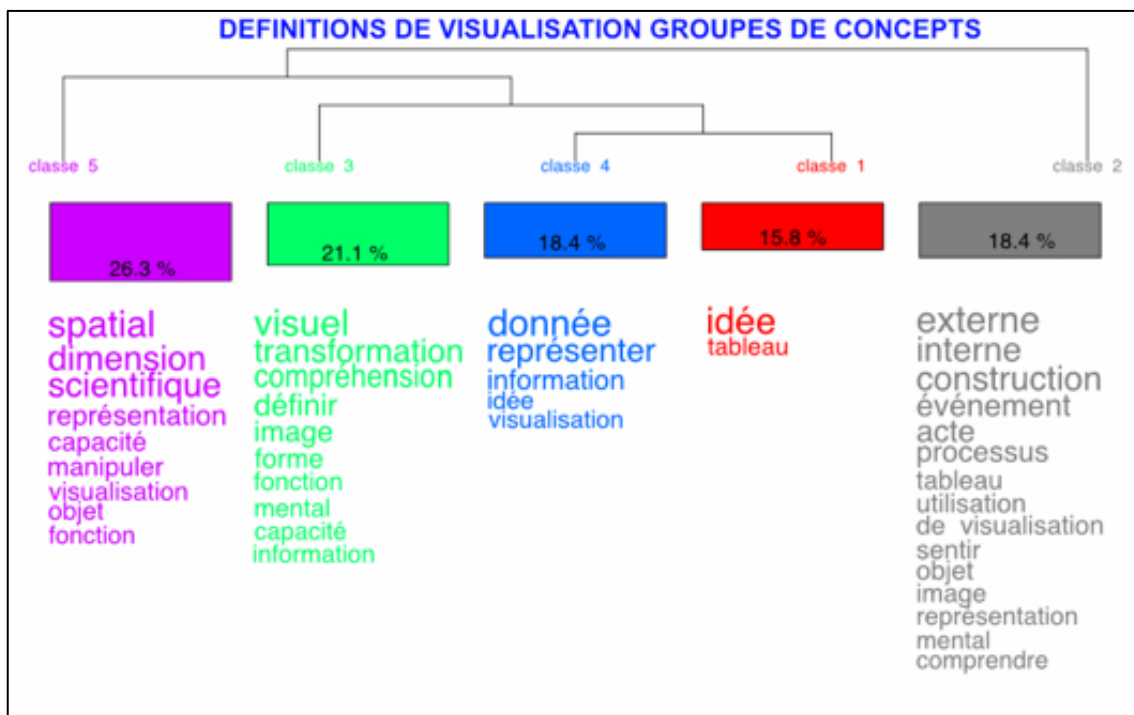


Figure 3. Dendrogramme de clusters de mots des définitions de la visualisation

¹² 1883 ds NED; v. Rey-Gagnon Anglic. 1981. Bbg. Quem. DDL t. 10. « VISUALISATION n. f., emprunté (1887) à l'anglais *visualization* (1883), de "*to visualize*". (ALAIN, 1992)

¹³ Voir le site https://mschermann.github.io/data_viz_reader/introduction.html#what-is-data-visualization

II. La visualisation des données, de l'information et des connaissances

Les étapes du processus de la visualisation

On peut identifier six étapes différentes dans le processus de la visualisation (Chitarrò, 2004) : la cartographie, la sélection, la présentation, l'interactivité, la convivialité et l'évaluation. Pour savoir si la méthode de visualisation est efficace ou non, si le but est atteint ou non, une évaluation de la visualisation est nécessaire. L'évaluation de « l'utilisabilité »¹⁴ a été développée pour les interfaces graphiques, adaptée et étendue aux interfaces de visualisation d'information. On peut classer les méthodes d'évaluation en deux grandes catégories : l'évaluation analytique qui comprend les examens d'experts et l'étude cognitive. L'évaluation empirique comprend les expériences contrôlées, les questionnaires, les entretiens et les groupes de discussion (Chitarrò, 2006).

Les taxonomies de la visualisation

Plusieurs taxonomies ont été proposées qui s'appuient sur la nature des données ou sur les techniques de visualisation, ou le type de données ou bien sur le type d'interaction. (Solweig, 2006).



Figure 4. Les types de visualisation. Source <https://datavizcatalogue.com/index.html>

¹⁴ L'utilisabilité, ou encore aptitude à l'utilisation est définie par la norme ISO 9241-11 comme « le degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficacité et satisfaction, dans un contexte d'utilisation spécifié ».

III. La visualisation des données dans le droit

État de l'art de la visualisation des données jurisprudentielles

Le rapport présente un état de l'art de la visualisation de données jurisprudentielles. L'analyse et la visualisation du réseau des décisions judiciaires peuvent mettre en lumière les relations entre des personnes et des institutions qui créent ou apparaissent dans les sources du droit. Parmi les systèmes de visualisation de la jurisprudence qu'on peut trouver, nous analysons les plus importants. EuCaseNet, sur la jurisprudence européenne, permet d'appliquer les techniques de l'analyse de réseaux à l'ensemble du corpus de la jurisprudence de l'UE (arrêts, ordonnances, avis et conclusions de l'avocat général de la Cour de justice et du Tribunal, à partir de leur date de création) en format XML. Le moteur de recherche de la jurisprudence constitutionnelle du Tribunal constitutionnel espagnol est un exemple concret des possibilités offertes par la visualisation de la jurisprudence. Dans le site <https://hj.tribunalconstitucional.es/HJ/es/Busqueda/Index>, le système propose une visualisation de l'ontologie juridique. En cliquant dans chacun de ces 4 cercles, on va afficher une nouvelle visualisation de façon hiérarchisée des descripteurs qui font partie d'un thésaurus de la base de données des arrêts de la cour constitutionnelle.

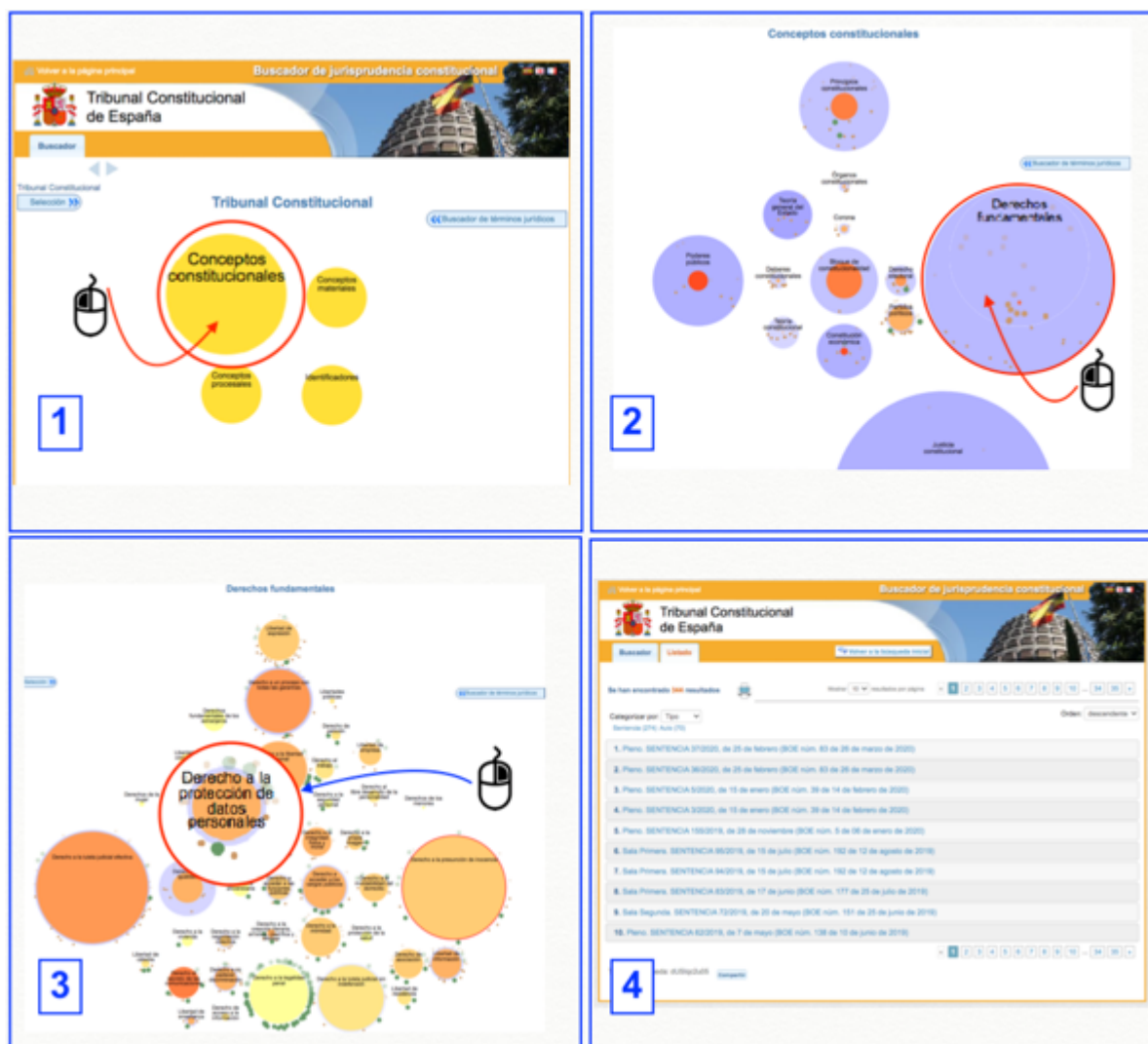


Figure 5: Visualisation de l'ontologie juridique de la Cour constitutionnelle d'Espagne

Des projets et produits de visualisation de la jurisprudence sont développés aux États-Unis et montrent l'importance de cette approche non seulement au niveau de la recherche universitaire, mais aussi des grandes sociétés de bases de données juridiques qui intègrent dans ces produits des modules de visualisation, en particulier de visualisation analytique, pour satisfaire la demande croissante des juristes et magistrats. Deux exemples sont Ravel Law et Court Listener. Ravel Law¹⁵ offre un accès à la recherche et à la visualisation juridique assistée par ordinateur. Ravel Law permet aux avocats de trouver, contextualiser et interpréter des informations et des données juridiques qui deviennent des connaissances juridiques. *Ravel Search Visualization* génère une carte de la jurisprudence pertinente pour aider les avocats à comprendre le paysage juridique et à s'assurer que les causes clés ne sont pas oubliées. L'autre exemple, Court Listener, est une base de données juridiques gérée par le « Free Law Project ».¹⁶ Court Listener est un site Web de recherche juridique gratuit contenant des millions d'avis juridiques provenant de tribunaux fédéraux et d'États. Avec Court Listener, les avocats, les journalistes, les chercheurs universitaires peuvent se tenir au courant des nouvelles informations au fur et à mesure qu'elles sont présentées, ou faire une analyse approfondie en utilisant les données du site. Les visualisations des décisions de la Cour suprême permettent de consulter et d'analyser les précédents de jurisprudence dans la base.

Casemine¹⁷ est une plateforme de recherche et d'analyse juridique en ligne qui offre une solution visuelle unique à tous les besoins de recherche juridique. Avec un groupe d'anciens élèves de l'IIT (M), de l'IIT (B) et de l'IIM (A) travaillant sous la direction de l'ancien juge en chef ainsi que de certains des meilleurs avocats indiens, CaseMine a réussi à cartographier la jurisprudence indienne dans une granularité jamais vue auparavant en combinant des avancées de pointe en intelligence artificielle et en visualisation avec des concepts de recherche juridique occidentaux modernes. La base de données jurisprudentielle indienne Manupatra¹⁸, propose plusieurs types de visualisation de la jurisprudence. On peut en retenir deux en particulier : une cartographie de la jurisprudence et une ligne de temps pour monter les citations les plus importantes.

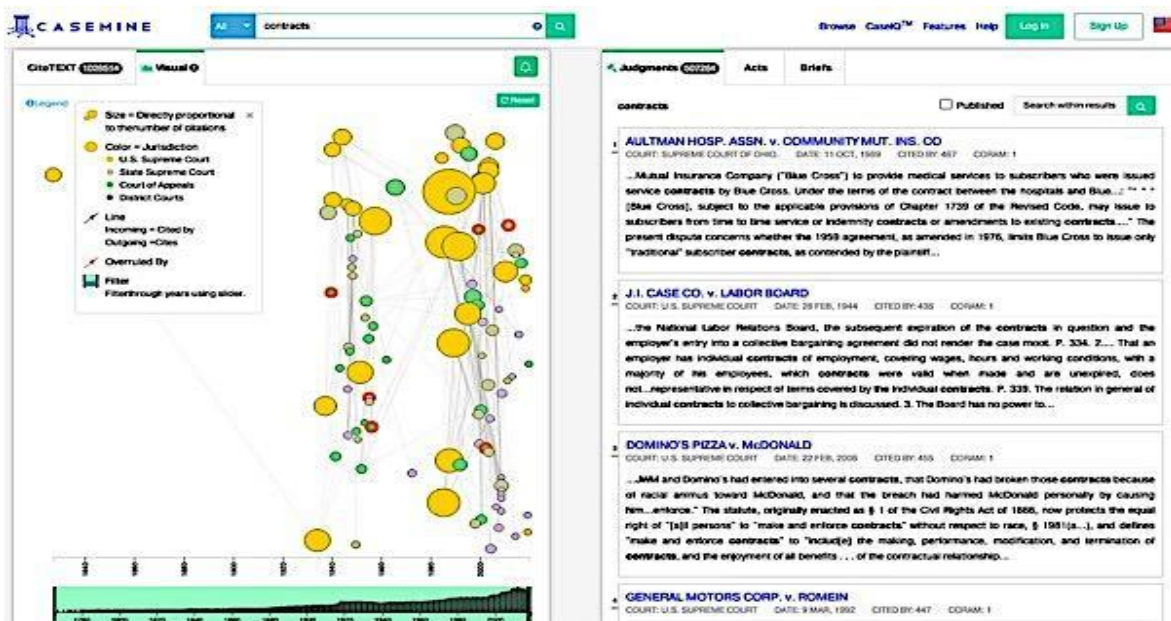


Figure 6: Plateforme de recherche et d'analyse juridique Casemine

¹⁵ Voir le site <http://ravellaw.com/>

¹⁶ <https://free.law/>

¹⁷ Voir le site <http://www.casemine.com>

¹⁸ Voir le site <https://www.manupatrafast.com/Home.aspx>

IV. Les fondements théoriques de la recherche

Le droit comme système complexe

Les hypothèses de recherche proposées par notre projet sont :

a) L'étude de la visualisation de systèmes complexes en droit⁴ pour valider : 1) l'organisation enchevêtrée des connaissances de cas de jurisprudence et 2) la connaissance des liens et relations qui relient ces composants et surtout les tendances d'évolutionnaires de la jurisprudence. Ces recherches doivent prendre en compte les caractéristiques fondamentales du droit : une connaissance auto-organisée et interconnectée suivant un modèle réticulaire du droit. (Bourcier, 2011)

b) L'application des méthodes de visualisation analytique de la jurisprudence (Lettieri, 2018). L'approche d'un tel système par sa complexité nous permet de concevoir les outils d'analyse visuelle de cette complexité à l'aide des analyses linguistiques. Dans notre projet, nous avons pu établir différents types de visualisations selon les ensembles de données traitées. De cette façon, on peut distinguer dans notre projet trois niveaux d'analyse (Börner, 2014) : niveau micro (individuel), qui regroupe un ensemble de données qui va jusqu'à 50.000 arrêts ; niveau méso ou méso-analyse (local) qui se fait avec un ensemble de données allant de 50.000 à 250.000 ; niveau macro ou macro-analyse, où les ensembles de données dépassent généralement 2.000.000 enregistrements.

Dans chaque niveau d'analyse, nous étudions cinq types de relations dans la jurisprudence : les relations statistiques/profilage, les relations temporelles, les relations sémantiques, les relations thématiques et relations de réseaux des acteurs.

c) La création de dispositifs de décision en droit intégrant une cartographie interactive visuelle. Ce traitement visuel des données juridiques vise à transformer des données textuelles de la jurisprudence en une cartographie lisible et compréhensible pour les praticiens et chercheurs en sciences juridiques (Arruabarrena, 2015).

Le projet de Visualisation Interactive de la jurisprudence de la COur de Cassation (VICO) est une représentation visuelle en réseaux du corpus jurisprudentiel (Lettieri, N et al. 2015) Open Data présent sur le site <https://www.data.gouv.fr/fr/datasets/cass/>. Près de 4 millions de décisions de jurisprudences sont rendues chaque année par les tribunaux français. Les textes de la jurisprudence sont des documents complexes structurés. La complexification du droit est un phénomène croissant. La compréhension des systèmes complexes passe nécessairement par des modélisations et l'utilisation d'outils de visualisation (Bourcier, 2011). Une variété de travaux théoriques et appliqués a tenté de mesurer la complexité de la loi. Du côté théorique, deux communautés intellectuelles ont manifesté un intérêt particulier pour la complexité juridique. Plus précisément, des chercheurs dans les sous-domaines du droit et de l'économie (White 1992 ; Kaplow 1995 ; Tullock 1995 ; Epstein 1995, 2004; Kades 1997; Frisch 2010) et de l'intelligence artificielle et (Bibel 2004; Pagallo 2010; Bourcier et Mazzega 2007) ont proposé des contributions à diverses dimensions pertinentes pour traiter cette question au sens large. L'approche d'un tel système par sa complexité nous permet alors, comme dans les autres disciplines, de développer les outils d'analyse scientifique et technique de cette complexité et les algorithmes de son usage.

Le point de départ de notre étude est la notion fonctionnelle du document juridique. Dans cette approche, le document a d'une part une fonction de communication, et d'autre part, une fonction de mémoire. Nous développons une approche tridimensionnelle du document juridique en tant que forme, contenu et médium, à partir des recherches menées par le collectif T. Pédaque (2006). Enfin, l'analyse de l'information jurisprudentielle passe par des opérations communes aux linguistes et aux informaticiens : analyse lexicale, syntaxique et sémantique.

Open data juridique-BDJ-Bases jurisprudentielles

L'hypothèse de départ de notre travail c'est aussi une approche pragmatique des données juridiques ouvertes, en particulier, des "données jurisprudentielles ouvertes". Pour s'engager dans cette approche, nous abordons d'abord la notion d'Open Data *stricto sensu*, entendue comme « données publiques ou privées organisées en bases de données, librement téléchargeables et réutilisables sous licence d'exploitation sans aucune contrepartie financière », puis la notion d'Open Data *lato sensu*, comprise comme " le traitement et l'analyse des Open Data par différentes techniques : statistiques, probabilités, data mining et apprentissage automatique ". L'Open Legal Data s'inscrit dans le cadre du grand projet gouvernemental français de libération des données numériques. Ces données peuvent être librement réutilisées pour développer des projets privés ou d'intérêt général avec quelques difficultés¹⁹. Nous démontrons que la visualisation est ici un outil fondamental.

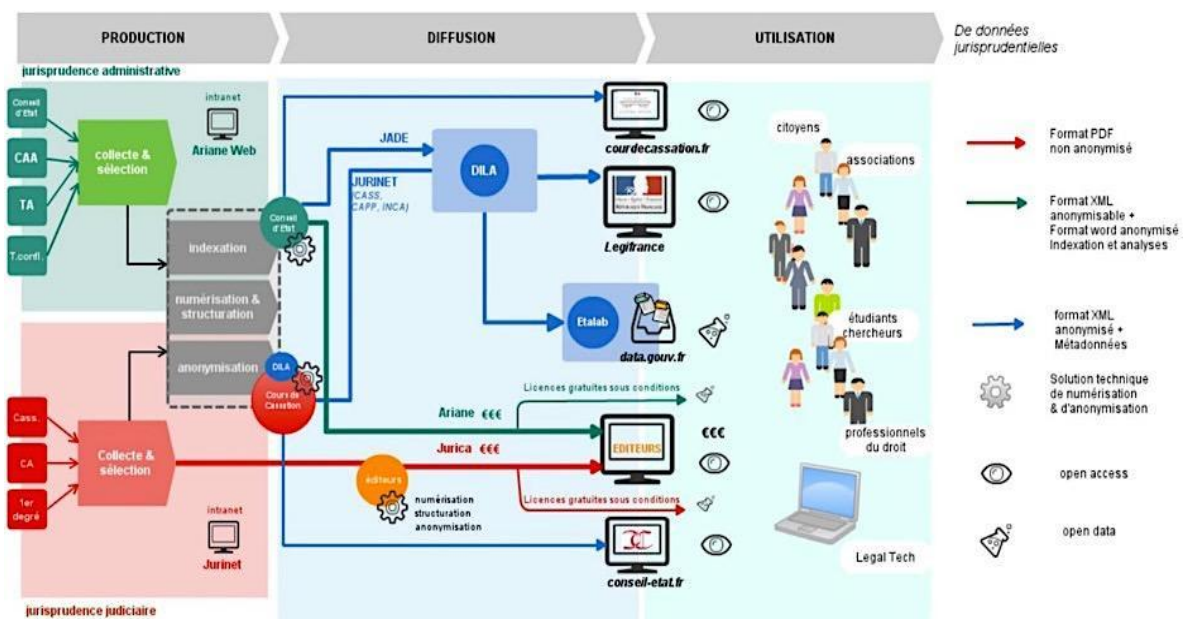


Figure 7: L'open data jurisprudentiel https://openlaw.fr/sites/default/files/2017-05/Livret_blanco_interactif21_04_0.pdf.
Licence CC SA 4.0)

Après avoir retracé l'historique et l'évolution des bases des données juridiques, nous faisons le point sur la jurisprudence d'appel qui est aujourd'hui classée, mais n'est pas toujours librement accessible sur internet. L'ordre judiciaire comprend des juridictions du fond, de première (TGI et TI) et seconde instance (cours d'appel), ainsi que de la Cour de cassation. Ces juridictions rendent des décisions de deux types : civiles et pénales. Les arrêts contenus dans la base Jurinet sont diffusés, après pseudo anonymisation, sur le site Légifrance et en Open data sur data.gouv.fr. La diffusion doit respecter les conditions applicables au service public de diffusion du droit par l'internet, comme le prévoit l'article R. 433-3 du code de l'organisation judiciaire ; la base CASS concerne les décisions « publiées » de la Cour de cassation ; la base INCA contient les décisions dites « inédites » de la Cour de cassation et la base CAPP, la sélection de décisions des juridictions du fond.

¹⁹ La jurisprudence en matière d'Open Data conduit également à des réflexions et des débats sur les problèmes posés par cette évolution, et concernant : a) Le cadre juridique ; b) l'anonymisation des décisions et c) la protection des données personnelles. La jurisprudence ouverte est à l'origine une source principale de développement car elle génère de l'innovation et crée une prévisibilité juridique.

V. Les fondements méthodologiques de la recherche

L'analyse de réseaux (ARS) en droit

Le projet VICO propose une application de la méthode des réseaux sociaux au domaine juridique. « *Le changement de paradigme dans les techniques de traitement des données qui déplace les techniques de visualisation du tableau de chiffres vers le graphe nœuds-lien est aussi le témoin d'une transformation des manières de construire les connaissances, en renvoyant aux aspects d'ingénierie documentaire, autant qu'aux aspects sémantiques et sémiotiques de la communication* » (Bourcier, Mazzega & Boulet 2010).

L'analyse des réseaux a permis de mettre en évidence un certain nombre de résultats qui apparaissent sur les relations qui se nouent entre décision de justice et le domaine du droit. Elle offre l'opportunité de décrire les réseaux structurés autour de cas de jurisprudence, de sujets qui génèrent le plus de litiges et la façon dont la Cour de Cassation décide sur les procès. Dans le domaine juridique, l'analyse de réseaux (ARS) favorise : a) de nouvelles approches techniques pour extraire, manipuler et analyser l'information juridique, b) la création de dispositifs sociotechniques visuels qui s'adressent à la fois aux juges, aux juristes et en général aux chercheurs en droit, sociologie, information et communication pour l'analyse des documents juridiques.

De nombreuses études ont montré la pertinence de modéliser les corpus juridiques comme des réseaux complexes. Un nouvel axe de recherche est apparu, qui consiste à considérer les décisions de justice comme des éléments de réseaux complexes, compréhensibles à partir d'une analyse structurelle des relations entre les décisions basées sur la théorie des graphes. Nous suivons le modèle de Jacques Bertin, c'est-à-dire, partir de l'analyse systématique de centaines de documents pour définir les variables visuelles et structurer les premières règles de construction du graphique en réseaux de la jurisprudence (Bertin 1967). L'Analyse des Réseaux Sociaux (ARS) mixe sociologie, statistique et théorie des graphes. L'analyse de réseaux se rapporte aux théories relationnelles qui permettent de formaliser les interactions en termes de nœuds et de liens. Après avoir retracé l'histoire de la théorie des graphes et les éléments fondamentaux de cette théorie, nous contextualisons l'ARS en droit.

L'analyse réticulaire, ou en réseaux, de la connaissance juridique, qui constitue une autre forme de modélisation sous forme de graphes, intègre la « modalité » ou les agrégations sporadiques de relations juridiques et s'offre comme une nouvelle dimension de la science juridique. Le défi actuel pour un juriste réside dans sa capacité à faire preuve de cohérence et d'intelligibilité de la complexité de ces différents réseaux et normes, constitués par des éléments disparates. L'analyse de réseau est largement utilisée dans divers domaines de recherche du droit, et particulièrement dans l'analyse de la loi et dans la jurisprudence (Bommarito, 2010). C'est une tendance qui a été détectée par des chercheurs depuis des décennies (par exemple, Merkl et Schweighofer, 1997). Des publications récentes telles que des conférences scientifiques Jurix (Boella, 2014 ; Hoekstra, 2014) et ICAIL (Icail, 2015) attestent de l'importance des structures de réseau dans la discipline juridique.

L'ARS est utilisé à plusieurs niveaux : pour analyser qualitativement et quantitativement les réseaux de citations juridiques (Agnoloni, 2015), pour sensibiliser à la complexité des textes juridiques (Bommarito, 2010 ; Walt, 2014), ou pour construire des systèmes de recommandation sur les bases de données d'information juridique (Winkels, 2014). L'utilisation de l'analyse des réseaux sociaux en droit fait l'objet d'études et de pratiques en Europe et aux États-Unis. Les travaux théoriques (Boulet, 2010 ; Bourcier, 2010 ; Bommarito, 2009, 2010 entre autres) montrent l'intérêt de cette méthodologie dans le domaine juridique. Nous établissons une synthèse exhaustive des sujets traités par divers auteurs autour du sujet de l'analyse de réseaux appliqué au droit.

VI. Résultats et livrables de VICO

Le processus de visualisation de la jurisprudence

VICO est un projet de recherche-action car d'une part, il développe des recherches de visualisation analytique (Visual Analytics) dans le domaine des sciences juridiques computationnelles et d'autre part, VICO est un dispositif intégrant un système visuel de recherche. Il propose un modèle conceptuel capable de prendre en compte les relations thématiques, temporelles et citationnelles dans un corpus jurisprudentiel, en s'appuyant sur des méthodes, processus et techniques de visual analytics (Text Mining, Data Mining, Big Data, Dataviz et Infoviz). VICO aborde la complexité de la connaissance juridique et la croissance exponentielle de la jurisprudence à l'aide de graphiques interactifs. Le travail porte sur la chaîne de valeur des données jurisprudentielles Open data : l'organisation et le traitement des arrêts et décisions dans des bases textuelles, l'indexation et l'analyse sémantique en s'appuyant sur la Science des données - data science - pour l'extraction, le traitement, la représentation et l'analyse visuelle des connaissances juridiques. Le traitement textuel du Big data juridique (BDJ) nous permet de créer un dispositif informationnel visuel à forte valeur ajoutée qui est supporté par une méthodologie en trois étapes :

1. Création de bases de données orientées graphe (Neo4J, Word Vectors 2 et Elasticsearch)
2. Modélisation de la connaissance à partir du datamining et text mining²⁰
3. Interfaces interactives et visuelles (Boella, 2014).

La visualisation interactive de la jurisprudence suppose comme on l'a vu, l'Open data juridique. Le projet de recherche s'inscrit dans la dynamique d'ouverture des bases de données jurisprudentielles (BDJ) impulsée par la Loi pour une République Numérique, prévoyant que « *les décisions rendues par les juridictions judiciaires sont mises à la disposition du public à titre gratuit dans le respect de la vie privée des personnes concernées* » (article L. 111-13 Code de l'Organisation Judiciaire). Le site <http://www.data.gouv.fr/fr/terms/> précise les conditions générales d'utilisation.

Pour la réalisation de notre projet VICO, nous avons dans un premier temps extrait du répertoire <ftp://echanges.dila.gouv.fr/CASS/> la totalité des fichiers XML qui présentent une structure qui permet de créer une base des données. Dans cette structure nous avons sélectionné les données et métadonnées juridiques pour créer notre base de données relationnelle. L'interface de la base de données présente tous les éléments et balises du document XML. Les fonctions relationnelles vont nous permettre de créer les relations entre les différents champs pour obtenir la visualisation.

Le projet VICO a été réalisé en deux (2) phases avec les outils suivants :

- Outil Vico—> Visualisation de la jurisprudence du site de la Cour de cassation. Système interactif visuel de la jurisprudence de la Cour de Cassation comportant les décisions de justice depuis l'année 2000 jusqu'à 2020 présentes sur le site www.cassation.gouv.fr avec une méthodologie d'analyse SNA (system networks analyse)
- Outil VisuaLEX—> Visualisation de la jurisprudence. Système utilisant les algorithmes de visualisation et la base de données de la jurisprudence de la Base de données CASS (datagouv.fr/cass) dans le cadre du projet Open data Jurisprudentiel (DILA)

Les thématiques, créées à partir des mots-clés de la jurisprudence ont été utilisées pour l'analyse micro et mezzo des rubriques établies pour la jurisprudence du site de la Cour de Cassation. Les rubriques du site de la Cour ont fait l'objet d'un classement hiérarchique car aujourd'hui, sur le site de la Cour, elles sont établies seulement à un niveau. Concernant la macro-analyse de la base CASS, nous sommes partis d'une part, de l'étiquette « RÉSUMÉ » contenue dans les fichiers XML de chaque décision de justice et d'autre part, de la table de rubriques du site de la Cour de Cassation.

²⁰ Le datamining et le text mining appliqués à la jurisprudence sont un ensemble de techniques qui allient les domaines de la linguistique, de la sémantique et du langage, des statistiques et de l'informatique. Ces techniques permettent d'extraire et de recréer de l'information à partir d'un corpus de textes.

Le projet VICO s'articule autour de 4 phases majeures du cycle de la visualisation des données jurisprudentielles : la première tourne naturellement autour de la collecte et des statistiques. Vient ensuite le « nettoyage des données » dont il s'agit de vérifier la valeur, de les structurer, de les organiser. Avec « l'analyse des données » on définit le type de visualisation et les techniques à déployer pour assurer par exemple l'interactivité. Le dernier stade est celui de l'élaboration de la visualisation proprement dite. Les graphiques des réseaux du projet VICO montreront les analyses des thématiques les plus traitées dans la jurisprudence, les types de décision de la Cour, les décisions de la Cour d'Appel objet du recours, les références aux textes légaux. L'intention de notre travail théorique et pragmatique de la visualisation est de bénéficier des différents avantages des méthodes qualitatives (différents niveaux de profondeur des catégories juridiques) et quantitatives (taille de la base de jurisprudence, nombre de clusters thématiques, tendances dans les sujets des arrêts, généralisation possible de certaines relations).

La cartographie globale de la jurisprudence de la Cour de Cassation

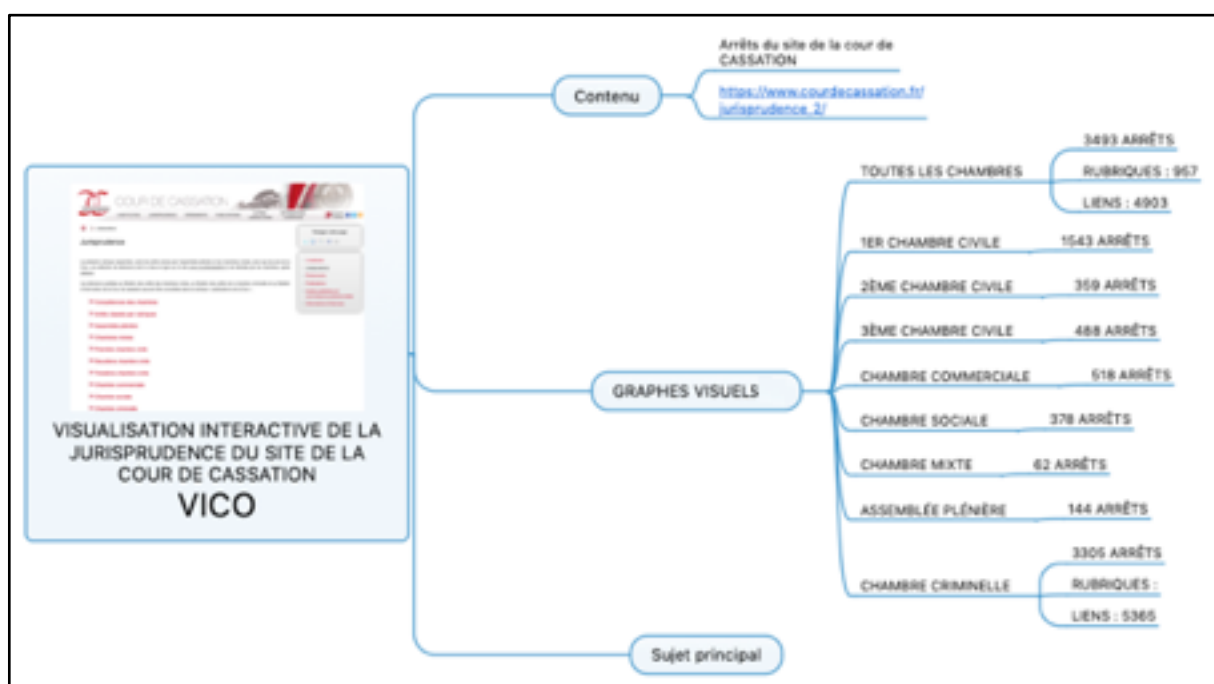


Figure 8. Carte conceptuelle de VICO.

VICO, pour la visualisation, prend seulement en compte les rubriques (titrages) disponibles dans le site de la Cour de cassation (une ou deux rubriques). En effet, l'absence d'autres titrages présents dans l'analyse faite par les magistrats, dans chaque jurisprudence a été prise en compte pour la réalisation avec Visualex, qui sera un système plus abouti. L'ensemble de données des 20 dernières années donnent le résultat suivant : 38.000 arrêts toute formation confondue. 475.000 descripteurs et 924 RUBRIQUES.

Ce travail a été présenté à la Cour de cassation au mois de juin 2019. De cette rencontre avec des utilisateurs mais aussi avec le responsable de contenu de la jurisprudence du site de la Cour de cassation, sont ressorties certaines observations, à savoir :

- Le projet de visualisation peut représenter une bonne alternative pour rechercher la jurisprudence de la Cour de cassation.
- Dans la visualisation, il faut distinguer les deux types de rubriques qui appartiennent à la Chambre criminelle et les rubriques des autres chambres.

- L'apport majeur constitue le fait d'ajouter des renseignements importants comme par exemple le champ ANA (Analyse du magistrat) ou les relations avec les liens de l'arrêt.

Nous avons pris note de ces recommandations et nous avons décliné de façon différente notre travail de visualisation. Une première chose à faire a été le fait de séparer les données de jurisprudence de la chambre criminelle, mais aussi de proposer une visualisation pour chaque chambre de la cour de cassation ainsi que de la chambre plénière et la chambre mixte.

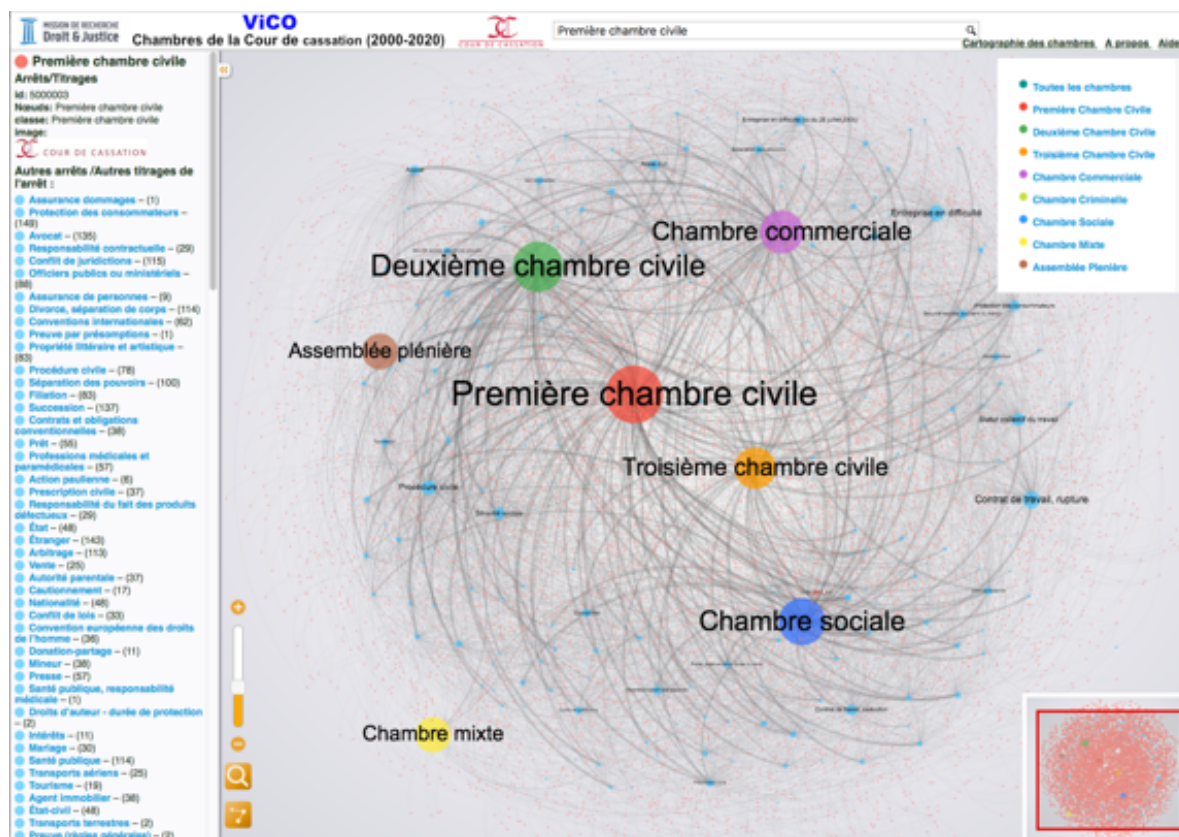


Figure 9. Visualisation Interactive par rubriques de la jurisprudence de la Cour de Cassation (VICO)
http://visualex.org/AC_WEB/AuCh_WEB.js/

La visualisation de la jurisprudence des bases CASS

Un exemple de la visualisation de ce système est le graphe de réseaux de la Chambre criminelle de tous les arrêts depuis l'année 2000 jusqu'à 2020.

La base CASS compte aujourd'hui 153.000 arrêts. Plusieurs contraintes ont fait que ce travail nous a pris beaucoup de temps :

- Dans 5% des arrêts, les données sont incomplètes.
- 30% des arrêts ont des problèmes de mise en format : double espace, rubriques collées, erreurs de frappe,
- Les données de la base CASS présentent beaucoup d'erreurs au niveau de l'organisation des rubriques (écrites en majuscules dans la base CASS) et de mots clés ou descripteurs (écrits en minuscules dans la base CASS)

Un traitement des données quasi manuel a été nécessaire pour organiser les données de la façon suivante :

Nous avons exporté les données dans le logiciel Open Refine pour les nettoyer. Nous avons procédé dans un premier temps à la correction des titrages (certaines décisions ont jusqu'à 32

rubriques). Ensuite nous avons séparé les titrages et leurs descripteurs pour leur assigner un identifiant, un n° de position et le numéro de la décision.

Ce travail va nous permettre d'établir les liens pour la visualisation du réseau entre le n° d'arrêt et les différents titrages, et de pouvoir ainsi reconstruire dans la base de données les relations de titrages dans la base CASS de manière automatique.

L'intérêt de cette démarche est de pouvoir établir les liens entre toutes les rubriques de tous les arrêts et pouvoir filtrer par niveaux, par RUBRIQUE et par formation (Chambre civile, commerciale, etc.)

https://www.legifrance.gouv.fr/juri/id/JURITEXT000039099254?dateDecision=&isAdvancedResult=&isAdvancedResult=&page=80&pageSize=100&query=* &searchField=ALL&searchProximity=&searchType=ALL&sortValue=DATE_DESC&tab_selection=juri&tab_selection=juri&typePaging=DEFAULT

Un exemple de la visualisation de ce système est le graphe de réseaux de la Chambre criminelle de tous les arrêts depuis l'année 2000 jusqu'à 2020. L'ensemble de données de 20 dernières années donnent le résultat suivant : 38.000 arrêts toute formation confondue. 475.000 descripteurs et 924 RUBRIQUES.

Le système interactive Visualex. <http://www.visualex.fr/>

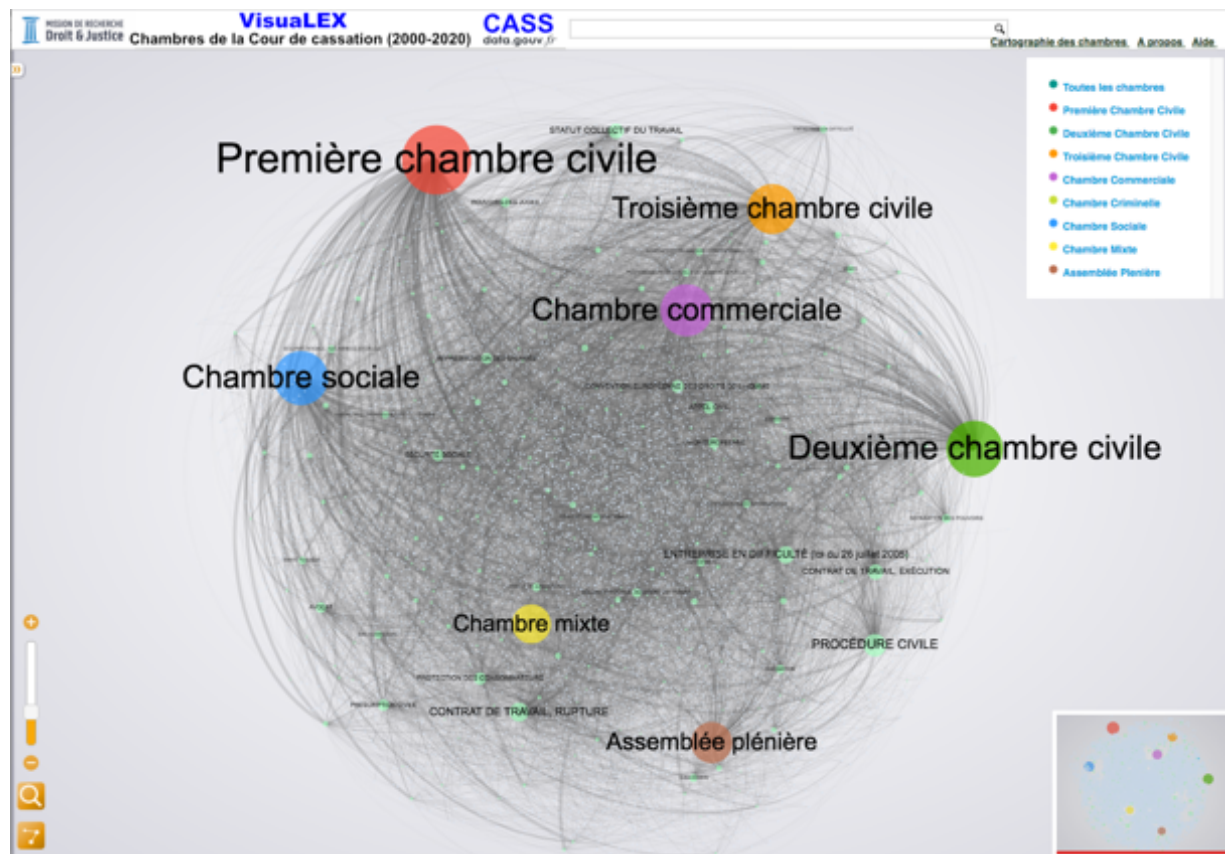


Figure 10. Visualisation Interactive de la jurisprudence de la base de données CASS

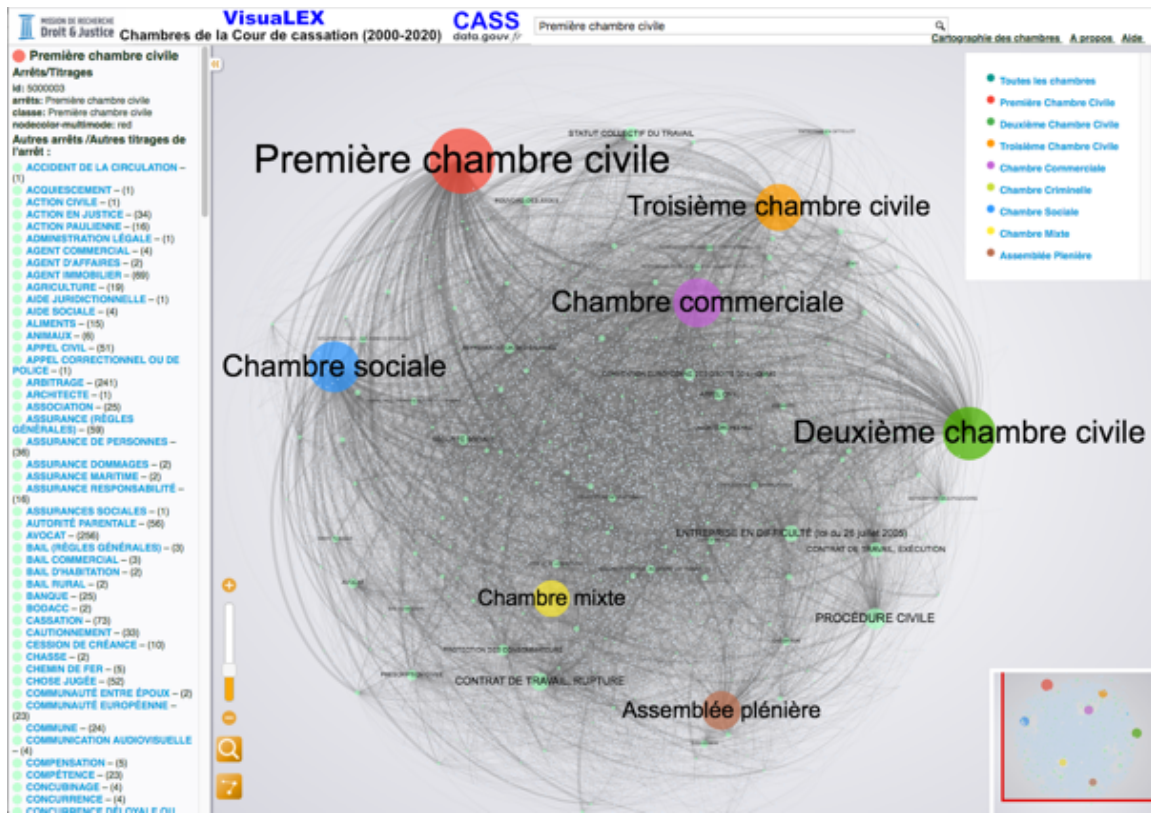


Figure 11. Visualisation Interactive par rubriques de la jurisprudence de la base de données CASS

VII. Conclusion

Notre travail de recherche sur la visualisation de la jurisprudence est à notre connaissance une première dans le domaine juridique français. C'est un projet dit de recherche-action, car d'une part il développe des recherches de visualisation analytique dans le domaine des sciences juridiques computationnelles, et d'autre part il requiert le traitement des données open data et la création d'un dispositif de visualisation interactive. Cette recherche-action dans le sens recherche pour l'action, contextuelle et par nature exploratoire, est basée sur des données retravaillées, contextualisées et analysées ce qui demande du temps et des méthodologies dites agiles.

Notre démarche a été d'abord de définir le cadre de notre recherche en partant des différentes définitions de la visualisation mais aussi en prenant compte de son évolution historique. Ainsi, on a pu observer que la présence des images et des graphiques élaborés avait été utilisées dans les manuscrits juridiques depuis le Moyen Age. Ceci nous a donné un présupposé valide pour montrer que la culture visuelle fait partie de la connaissance juridique et que cette approche pourrait être valable de nos jours, car avec le big data législatif et jurisprudentiel cette représentation visuelle pourrait aider à la compréhension de la complexité de la jurisprudence.

Notre recherche a montré qu'aujourd'hui les systèmes de visualisations ne représentent pas seulement des approches théoriques, mais sont d'importance dans la plupart de pays avec une tradition juridique importante. Les exemples de Ravel Law, de Casemine, de Lexisnexis montrent que l'apport de la visualisation dans le travail du juriste, du magistrat et de l'étudiant en droit permet de mieux appréhender la connaissance quand il s'agit de corpus assez vastes. Cependant, l'état de l'art de l'existant montre que la plupart des études se concentrent sur la Common Law²¹ : les juges à travers les décisions des tribunaux, qui est fondamentalement différente du droit civil qui est utilisé dans le

²¹ Les réseaux de décisions jurisprudentielles dans le Common Law contiennent par ailleurs des renseignements, capables d'identifier les précédents faisant autorité, d'évaluer la pertinence des décisions judiciaires ou même de prédire les causes qui seront citées davantage dans l'avenir.

droit continental. Ainsi, pour construire un modèle de réseau qui représente la jurisprudence de manière assez fiable, nous devons prendre en compte la structure multi-échelle du droit. Des traits distinctifs du droit, comme la hiérarchie entre les sources du droit ou les différents types de relations entre les documents juridiques, devraient être bien découpés et incorporés dans un modèle, comme nous le verrons plus en détail dans la suite du travail.

Pour atteindre les objectifs et en appliquer la méthodologie proposée, VICO propose une cartographie interactive de la jurisprudence française et fournit une réponse claire et synthétique à l'utilisation des données juridiques initialement complexes pour l'usager des données juridiques. La visualisation est un bon outil au service de la complexité pour avoir accès à l'hyperstructure de la jurisprudence constitutivement interconnectée et ceci depuis des données liées simplement les unes aux autres (Bommarito; Katz et Kelner, 2009). Dans notre projet, la visualisation a essentiellement pour but d'exprimer de façon claire et synthétique un problème initialement complexe et difficile à appréhender dans son intégralité. La visualisation doit donc mettre en valeur les éléments pertinents pour la décision jurisprudentielle à prendre, et leur donner du sens pour faire apparaître de façon claire et compréhensible une structure juridique souvent sous-jacente ou masquée au sein de données nombreuses et complexes.

VICO et VISUALEX sont des dispositifs qui répondent, sur le plan scientifique, à l'intégration toujours plus étroite entre sciences sociales et technologies de l'information dans le domaine émergent interdisciplinaire des Sciences juridiques computationnelles autour de la visualisation analytique (Dhruv, 2011). C'est un outil original et puissant de médiation au service des interactions Homme-données, car la visualisation analytique et l'aide à la décision donnent une valeur ajoutée aux informations jurisprudentielles destinées aux acteurs de droit (praticiens et chercheurs). Enfin, il faut préciser que quelles que soit la ou les méthodes utilisées, le rôle du juriste reste essentiel et ce principalement concernant la validation de l'analyse du corpus jurisprudentiel. Il ne s'agit en aucun cas d'un usage aveugle, mais au contraire d'être capable d'allier différentes méthodes, d'aiguiller les logiciels et autres algorithmes afin d'arriver, ensemble, à une analyse la plus complète et satisfaisante de la jurisprudence de la Cour de Cassation.