



HAL
open science

L'intelligence artificielle dans le domaine de la justice : réflexions sur la “ représentation sémantique ” du droit

Margo Bernelin

► **To cite this version:**

Margo Bernelin. L'intelligence artificielle dans le domaine de la justice : réflexions sur la “ représentation sémantique ” du droit. Penser, calculer, délibérer (Ph. PEDROT & Alain PAPAUX dir.), Mare & Martin, 2022, 978-2-84934-619-8. hal-03928199

HAL Id: hal-03928199

<https://hal.science/hal-03928199v1>

Submitted on 25 Jul 2024

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.

L'intelligence artificielle dans le domaine de la justice : réflexions sur la « représentation sémantique » du droit¹

Pour l'ouvrage *Penser, calculer, délibérer* – sous la direction de Philippe Pédrot – Mare & Martin – 2022

« Actuellement plusieurs entreprises innovantes du secteur juridique, souvent appelées Legal Tech et certains États proposent des services utilisant de l'intelligence artificielle pour améliorer et personnaliser les résultats de recherche sur des bases de données juridiques et pour faciliter la prise de décision notamment en matière de provision dans un contexte contentieux. Les dispositifs les plus ambitieux proposent même de calculer les risques de récidives, même si certains de ces outils déjà expérimentés ont parfois été encartés pour des raisons d'éthiques mais également d'inefficacité, comme dans le cas du système COMPAS aux États-Unis »². Prenant acte de cet engouement pour l'intelligence artificielle (IA) dans le champ de la justice et des difficultés soulevées, le Conseil de l'Europe étudie la possibilité de mettre en place un mécanisme de certification lequel permettrait d'assurer un certain contrôle des outils tout en suscitant la confiance des utilisateurs³. En France, le projet DataJust est emblématique tant de l'enthousiasme que des craintes soulevées par le déploiement de l'IA dans ce champ⁴. Le projet a pour ambition le développement d'un algorithme qui permettrait d'informer les juges, mais aussi les justiciables, sur les dommages et intérêts alloués en cas de préjudice corporel. Pour ce faire, « les montants demandés et offerts par les parties aux instances, les évaluations proposées dans le cadre de procédures de règlement amiable des litiges et les montants alloués aux victimes par les juridictions »⁵ seront recensés et les données pertinentes extraites des « décisions de justice rendues en appel entre le 1er janvier 2017 et le 31 décembre 2019 par les juridictions administratives et les formations civiles des juridictions judiciaires »⁶. Dès sa publication, le décret relatif au projet a fait l'objet de nombreux commentaires⁷.

L'effervescence suscitée par DataJust, dont la mise en œuvre démarre timidement, tranche avec le peu d'intérêt pour les éléments techniques qui sous-tendent ce type d'algorithmes. Ainsi,

¹ Margo BERNELIN, CR, UMR 6297 Droit et changement social (Université de Nantes).

² Conseil de l'Europe, *Mise en place éventuelle d'un mécanisme de certification des outils et services d'intelligence artificielle dans le domaine juridique et judiciaire, étude de faisabilité*, 8 décembre 2020, CEPEJ(2020)15 Rev.

³ < <https://rm.coe.int/draft-agenda-23sept2019-certification-work-2/1680973af5> >, dernier accès le 25 mars 2021.

⁴ Décret n° 2020-356 du 27 mars 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé « DataJust », JORF n° 0077 du 29 mars 2020 ; CNIL, Délibération n° 2020-002 du 9 janvier 2020 portant avis sur un projet de décret en Conseil d'État portant création d'un traitement automatisé de données à caractère personnel dénommé « DataJust » (demande d'avis n° 19020148) ; Décret n° 2020-797 du 29 juin 2020 relatif à la mise à la disposition du public des décisions des juridictions judiciaires et administratives.

⁵ Voir la présentation du projet sur le site du ministère de la justice, < <https://www.justice.fr/donnees-personnelles/datajust> >, dernier accès le 25 mars 2021.

⁶ Article 2 du décret n° 2020-356 du 27 mars 2020.

⁷ Alexandra BENSAMOUN et Thibault DOUVILLE, « Datajust, une contribution à la transformation numérique de la justice », *JCP G.*, 2020, n°19, pp. 907-910 ; Pierre JANUEL, « Datajust : un algorithme pour évaluer les préjudices corporels », *Dalloz actualité*, 01 avril 2020 ; Cécile CRICHTON, « Création de "DataJust", un algorithme prédictif d'évaluation du préjudice », *Dalloz IP/IT*, 2020 p.209 ; Eugénie PETITPREZ et Rodolphe BIGOT, « Standard humain ou standardisation algorithmique de l'évaluation du dommage corporel ? », *LexBase Pro*, actes de séminaire, 2021, n°310 ; Yannick MENECEUR, « DataJust, face aux défis de l'intelligence artificielle », *JCP G.*, 2020, n°40, pp. 1708-1714 ; Motahareh FATHISALOUT BOLLON et Vincent RIVOLLIER, « À propos de DataJust : justesse de l'outil numérique, juste indemnisation des victimes ? », *Revue Lamy Droit Civil*, 2020, n° 184, pp.18-24 ; Cédric COULON, « Datajust, ou l'outil référentiel numérique des préjudices corporels : aussitôt né, aussitôt enterré ? », *Responsabilité civile et assurances*, 2020, n°5, pp.3-4 ; Julien BOURDOISEAU, « "DataJust" ou la réforme du droit de la responsabilité civile à la découpe ? », *LexBase Hebdo*, 2020, n°821 ; Georges TEBOUL, « La justice prédictive : une actualité inquiétante, ou un pari exaltant ? », *La Gazette du Palais*, 2020, n°14, p.12.

rare sont les travaux des juristes sur les manières d'extraire les données, de les cataloguer, de les agréger, autant d'éléments qui auront pourtant une influence déterminante sur les résultats de l'algorithme. Dans cette perspective, si la face émergée de l'iceberg a fait l'objet de nombreuses recherches sous la bannière de la justice prédictive⁸, la face immergée reste, quant à elle, quasi-inconnue, laissée aux seuls initiés⁹. En témoigne, le manque de relais de la grande messe européenne ENDORSE (16-19 mars 2021)¹⁰, conférence surdimensionnée dédiée au référencement des données et à la sémantique, dont une journée était dédiée aux manières d'accéder et d'extraire les données dans le champ du droit. Cette extraction repose sur des outils dits sémantiques qui permettent à la fois d'identifier un document mais aussi sa thématique, voir les éléments pertinents de son contenu. Pour ce faire des mots clefs, prédéfinis et hiérarchisés entre eux, sont utilisés pour annoter les textes. Il s'agira d'indiquer que le texte est une loi, qu'il transpose une directive de l'Union européenne, de préciser sa date d'entrée en vigueur ou encore sa thématique. L'algorithme d'extraction des données n'aura pas à consulter le document dans son intégralité pour décider de sa pertinence, mais simplement à interroger les mots clefs, aussi nommés métadonnées. Ces dernières sont consignées dans des documents intitulés en science de l'information *ontologies* et permettent également de formuler des réponses face aux requêtes en s'appuyant sur les définitions fournies. Le vocabulaire pourra alors être utilisé pour indiquer au requêteur qu'il/elle consulte une loi relative à la responsabilité pénale et cela à l'appui de la définition de ce terme.

Le vocabulaire d'annotation facilite donc le travail des algorithmes tout comme il le contraint et conditionne les résultats offerts. La centralité de ces outils sémantiques impose dès lors d'en analyser la sélection, la construction mais aussi leurs finalités et particularités au regard du champ qu'est le droit. En science de l'information ces outils sont présentés comme permettant de « représenter un champ de savoir », c'est-à-dire d'en offrir une description ordonnée¹¹. Ils questionnent le juriste : le droit constitue-t-il un champ « représentable » sémantiquement ? À l'analyse, le droit ne se laisse pas facilement saisir par ces outils, la représentation du droit étant une ambition difficile à atteindre eu égard aux complexités et nuances qui traversent la matière (I). La représentation du droit par des outils sémantiques peut même s'avérer trop ardue, posant la question de la pertinence même de ces derniers pour décrire le droit. En particulier, le vocabulaire d'annotation retenu ainsi que la formulation très ouverte des définitions contrastent avec la spécificité et précision de la terminologie juridique remettant en cause la possibilité même d'une représentation sémantique du droit (II).

I) La représentation du droit : une ambition difficile à atteindre

⁸ Ordre des avocats au Conseil d'État et à la Cour de cassation, *La justice prédictive, Actes du colloque du 12 Février 2018*, Dalloz, Paris, 2018 ; n° spécial « Justice prédictive », *Archives de Philosophie du Droit*, Tome 60, 2018 ; Stéphanie LACOUR et Daniela PIANA, « Faites entrer les algorithmes ! Regards critiques sur la "Justice prédictive" », *Cités*, 2019, n°4, pp. 47-60 ; Antoine GARAPON et Jean LASSEGUE, *Justice Digitale*, PUF, Paris, 2018 ; Voir également le dossier « Les défis de la justice numérique », *Les Cahiers de la Justice*, 2019, n°2. Pour aller plus loin sur ce thème voir : Sonia DESMOULIN-CANSELIER et Daniel LE METAYER, *Décider avec les algorithmes : Quelle place pour l'Homme, quelle place pour le droit ?*, Dalloz, Paris, 2020 ;

⁹ Les travaux dédiés aux outils sémantiques sont rares en langue française mais l'on peut noter la thèse de Jean Leclercq, *Les représentations informatiques des connaissances juridiques : l'expérience française*, Université Lille II, 1999. Du côté de la doctrine étrangère, les travaux sont légèrement plus nombreux, voir notamment V. Richard BENJAMINS, Pompeu CASANOVAS, Joost BREUKER et Aldo GANGEMI (dir.), *Law and the Semantic Web (Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications)*, Springer, Berlin, 2005 ; Monica PALMIRANI, Ugo PAGALLO, Pompeu CASANOVAS et Giovanni SARTOR (dir.), *AI Approaches to the Complexity of Legal Systems (Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents)*, Springer, Berlin, 2012.

¹⁰ < <https://op.europa.eu/en/web/endorse> >, dernier accès le 20 mars 2021.

¹¹ En informatique le terme « représentation » est consacré mais son sens parfois discuté : voir Andre VALENTE, « Types and Roles of Legal Ontologies », dans *Law and the Semantic Web*, 2005, (précité) p. 66.

La numérisation du droit, son accessibilité en ligne ainsi que la volonté d'exploiter par des agents algorithmiques le contenu des règles publiées imposent de répertorier et classer les données. Cette étape est nécessaire pour opérer toute recherche, « fouille », sur ces données présentes en volume important et provenant de sources variées. En effet, mieux vaut entreprendre des recherches dans une bibliothèque bien rangée et disposant d'un catalogue consultable, que dans une salle où les livres ne sont ni répertoriés, ni classés. Dans le champ de l'informatique, les ontologies servent cet impératif de « rangement » et de « classement » des données en décrivant en amont les grandes caractéristiques utiles pour ce faire. Sorte de Vocabulaire juridique 3.0¹², les ontologies sont confrontées à un champ qui ne se laisse pas facilement saisir en raison de la complexité des éléments à représenter (**A**) mais aussi de leur caractère évolutif (**B**).

A) Le droit, un champ complexe

Tandis que certains éléments semblent facilement identifiables et représentables par les ontologies, d'autres apparaissent plus fuyants et délicats à saisir. L'ontologie *The Legal Knowledge Interchange Format Core (LKIF)*¹³ illustre ce contraste. Cette ontologie est l'une des plus volumineuses, comprenant 15 modules et des centaines de mots clefs. L'ambition de ce document est de fournir un vocabulaire « pour décrire des sets de concepts proches entre eux qui relèvent du champ du droit mais également du bon sens »¹⁴. Dans ce cadre, certains savoirs sont saisis de manière relativement fidèle par le vocabulaire proposé, à l'image du « code de conduite » qui est défini comme un « code décrivant les responsabilités ou les bonnes pratiques d'un individu ou d'une organisation, telle qu'une charte de comportements éthiques adoptée par une entreprise »¹⁵. L'ontologie précise que le terme « code de conduite » fait partie de la catégorie « Soft Law » (droit souple) laquelle est définie comme « faisant référence à un instrument quasi-légal qui n'a pas de force obligatoire ou qui détient une force obligatoire en quelque sorte plus faible que celle du droit traditionnel, souvent nommé "droit dur" dans ce contexte. Le terme "droit souple" est initialement apparu dans le champ du droit international, mais plus récemment il a été transféré dans d'autres branches du droit »¹⁶. Les définitions offertes sont ici pertinentes puisque l'on retrouve les caractéristiques essentielles des termes. En effet, du côté du droit souple, *LKIF* fait bien référence l'absence de contrainte tout comme à sa normativité « graduée » par rapport au droit dur¹⁷. La représentation du code de conduite est elle aussi adaptée : un document rassemblant les obligations et bonnes pratiques que se fixe une personne ou un groupe, bien que la doctrine juridique fasse plus souvent référence au groupe qu'à la personne physique. Ces éléments vont alors permettre de pouvoir fidèlement annoter des documents.

¹² Il s'agit ici d'un clin d'œil au *Vocabulaire Juridique*, ouvrage de référence en France. Gérard CORNU (dir.), *Vocabulaire Juridique*, Paris, PUF, 2020.

¹³ La rédaction de cette ontologie a été financée par le programme de l'Union européenne ESTRELLA (*European project for Standardised Transparent Representations in order to Extend Legal Accessibility*). Les membres du consortium ayant participé à la création de LKIF sont les suivants: Dó-és Pénzügyi Ellenőrzési Hivatal ; Budapesti Corvinus Egyetem ; Centro Nazionale per l'Informatica nella Pubblica Amministrazione ; Consorzio Pisa Ricerche SCARL ; Fraunhofer FOKUS ; Haley Limited ; Interaction Design ; KnowledgeTools ;

¹⁴ < http://www.estrellaproject.org/?page_id=3 >, dernier accès le 20 mars 2021.

¹⁵ « A code outlining the responsibilities of or best practice for an individual or organisation, such as a set of principles of good corporate behaviour adopted by a business », entrée « code of conduct », < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

¹⁶ « The term "soft law" refers to quasi-legal instruments which do not have any binding force, or whose binding force is somewhat "weaker" than the binding force of traditional law, often referred to as "hard law", in this context. The term "soft law" initially appeared in the area of international law, but more recently it has been transferred to other branches of law » ; entrée « soft law », < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

¹⁷ Pour une étude de référence sur la notion, voir l'étude du Conseil d'État, *Le droit souple*, Paris, La Documentation française, 2013.

Contrastant avec cet exercice réussi, d'autres termes sont eux plus complexes à saisir, les définitions offertes se trouvant soit approximatives, soit incomplètes. C'est le cas du mot « directive », défini par l'ontologie comme suit : « *Examples are European Union directive, a legislative act of the European Union and Directives, used by United States Government agencies (particularly the Department of Defense) to convey policies, responsibilities, and procedures* »¹⁸. La phrase est délicate à traduire eu égard à sa formulation maladroite, mais l'on peut proposer l'énoncé suivant : « Les exemples de directives sont les directives de l'Union européenne, actes législatifs de l'UE d'une part, et les directives des agences fédérales américaines (notamment celles du Département de la Défense) lesquelles permettent de communiquer des orientations politiques, des attributions de tâches ou encore les procédures à suivre, d'autre part ». Cette définition, qui tente à la fois de représenter le droit de l'Union et le droit des États-Unis, soulève des difficultés. Du côté du droit de l'Union, la définition est succincte et conduit à annoter tout texte législatif de « directive », alors même qu'il peut s'agir d'un Règlement. Par opposition, le site internet de l'Union européenne en propose une description plus complète, la directive étant un document qui « contraint les États membres quant aux résultats à atteindre mais laisse aux autorités nationales le choix des formes et des méthodes »¹⁹. Du côté du droit américain, le terme ne connaît pas de définition unique, certaines agences utilisant le mot pour leurs documents, d'autres préférant celui de « *regulation* » pour communiquer leurs orientations, rôles et procédures. Par ailleurs, le mot « directive » peut aussi correspondre à des documents non contraignants, de simples informations, ce que la définition de l'ontologie *LKIF* peine à traduire. Le terme « directive » apparaît donc complexe, faisant référence à réalités différentes selon les ordres juridiques, polysémie qui limite la représentation fidèle du mot « directive » en une seule catégorie.

Si la polysémie d'un terme rend ce dernier difficilement représentable, d'autres termes ne sont, eux, pas décrit de manière exhaustive, laissant des caractéristiques essentielles de côté. C'est le cas de l'entrée « coutume », définie par l'ontologie *LKIF* comme « un souvenir collectif d'un groupe, c'est-à-dire des propositions véhiculées par la coutume qui sont partagées (c'est-à-dire appliquées) par tous les membres du groupe »²⁰. Cette définition est elle-même complexe à saisir en raison de sa référence au souvenir et au fait que la précision définitionnelle soit apportée par l'usage du mot que l'on souhaite représenter (« véhiculée par la coutume »). Elle s'approche toutefois de ce que la doctrine entend : des règles reçues par un groupe et appliquées par ce dernier. Cependant, il manque un critère essentiel de la coutume : la constance de son application par le groupe. En droit international notamment cette caractéristique est centrale, ce que manque de présenter l'ontologie²¹. Plus complexe qu'il n'y paraît à représenter, le terme ne se laisse pas facilement saisir, ce qui est regrettable tant la coutume est une source de droit qu'il est crucial de repérer lors d'une recherche documentaire sur un thème.

Ces difficultés définitionnelles face à un champ qui ne se laisse pas aisément représenter ne sont pas exclusives à l'ontologie *LKIF* mais se retrouvent également au sein d'autres ontologies. C'est le cas de l'ontologie *GDPRtEXT*²², développée par des chercheurs du *Trinity College* (Dublin,

¹⁸ < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

¹⁹ « A directive is binding, as to the result to be achieved, upon each Member State to which it is addressed, but leaves to the national authorities the choice of form and methods », voir « DIR » < <https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/resource-type> >, dernier accès le 20 mars 2021.

²⁰ «The collective memory of some group of agents, i.e. propositions beared by custom are shared (i.e. held) by all members of the group », entrée « custom », < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

²¹ Entrée « coutume », Gérard CORNU (dir.), *Vocabulaire Juridique*, Paris, PUF, 2020.

²² Pour une présentation de l'ontologie voir Harshvardhan J. PANDIT, Kaniz FATEMA, Declan O'SULLIVAN, Dave LEWIS, « GDPRtEXT - GDPR as a Linked Data Resource », in *Proceedings of 15th European Semantic Web Conference (ESWC)*, Crete, Heraklion, Greece. 2018.

Irlande). Elle se veut être un outil de référence pour décrire le Règlement européen de protection des données (RGPD) en faisant la synthèse de ses 173 considérants et 99 articles²³ ! En pratique, le document offre les représentations des principales notions du RGPD et fait référence aux articles et considérants en lien avec elles. Pourtant ici aussi les représentations sont incomplètes tant les caractéristiques des termes sont nombreuses. C'est le cas de l'entrée « droit à l'effacement », décrit comme le droit pour la personne concernée par un traitement de données personnelles d'obtenir l'effacement de ses données personnelles²⁴. *GDPRtEXT* précise que ce droit est défini par l'article 17 du règlement, par les considérants 65, 66 et 68 et fait partie de la catégorie « droits des personnes concernées par le traitement » (concept racine ici). Le document opère un lien entre le droit à l'effacement et les entrées suivantes : « effacement si le consentement est retiré » et « effacement si cela n'est plus utile pour la finalité du traitement ». Ces deux sous-entrées reprennent les deux premières conditions de l'article 17(1) du RGPD qui permettent à un individu de demander l'effacement des données. En effet, l'article 17(1)(a) précise que le droit à l'effacement est disponible lorsque « les données à caractère personnel ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées ou traitées d'une autre manière » et le (b) lorsque « la personne concernée retire le consentement sur lequel est fondé le traitement ». L'ontologie résume donc à sa manière les caractéristiques et conditions du droit à l'effacement des données sans pour autant que ce résumé ne soit exhaustif. En effet, d'autres critères permettent à l'individu de demander l'effacement de ses données, notamment lorsque le traitement des données est illicite. Aussi toute annotation d'un texte grâce à cette référence au RGPD porterait le risque de manquer les autres conditions, non visées par l'ontologie et pourtant tout aussi essentielles.

Ces difficultés définitionnelles peuvent faire l'objet de corrections, mais elles témoignent de la nécessité de vérifier ces documents avant toute annotation. Dans cette perspective, la prise en compte de l'évolution du droit devra, elle aussi, faire l'objet d'une attention toute particulière.

B) Le droit, un champ évolutif

La représentation du droit grâce à des ontologies se heurte à une autre difficulté de taille : l'évolution du droit qui conduit à préférer des notions plutôt qu'à les remplacer ou à faire évoluer leur contenu rapidement. Cela impose aux auteurs d'amender si besoin l'ontologie en répertoriant les versions mises à jour du document. À titre d'illustration, l'ontologie *ELI (European Legislation Identifier)*, développée par l'Union européenne pour répertorier les textes de droit de l'Union et de ses États membres en est à sa troisième version²⁵. Les évolutions de l'ontologie sont précisément documentées, la version 1.3 publiée sous forme de tableur indique par surbrillance les entrées ayant été modifiées comme l'entrée « édicté par » (*passed by*) qui a évolué pour inclure dans la dernière version aussi bien les organismes qui créent du droit en l'édictant que les organismes intervenus dans la rédaction du droit qui a été édicté. Certaines ontologies dédiées à une branche du droit ou à un texte, aussi nommées *domain ontologies*²⁶, peuvent indiquer des concordances entre les normes décrites et celles qui les précédaient à l'image de l'ontologie *GDPRtEXT* qui offre une

²³ Règl. UE 2016/679 du 27 avr. 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, JO L 119 du 4 mai 2016.

²⁴ « The data subject has the right to obtain erasure of their personal data », < http://openscience.adaptcentre.ie/projects/GDPRtEXT/dpd_mapping.html >, dernier accès le 20 mars 2021.

²⁵ Cette ontologie comprend des entrées très générales, utiles pour répertorier les textes juridiques de l'Union et de ses États membres : « ressource légale », « juridiction », « date en vigueur », « date de publication », « consolide », « transpose » ou encore « amende ». Présentation d'ELI voir < http://publications.europa.eu/resource/cellar/917d16b5-28f6-11e8-b5fe-01aa75ed71a1.0001.01/DOC_1 > dernier accès le 20 mars 2021.

²⁶ Les informaticiens distinguent les « *core ontologies* » dédiées aux savoirs généraux et les « *domain ontologies* » qui s'attachent à des branches du droit ou à des corpus précis.

table de concordance entre le RGPD, décrit, et la directive de 1995 abrogée par le règlement²⁷. Ces éléments constituent un travail conséquent de veille juridique lequel ne doit pas être sous-estimé mais qui, hélas, ne fait pas toujours l'objet d'explicitations précises au sein d'autres ontologies.

Cette veille juridique peut constituer une tâche complexe lorsque les évolutions du droit sont initiées par les juges ou des agences ayant pour mission l'application des textes de droit. Ce sont alors les interprétations des dispositions qui enrichissent la règle édictée, ce qui impose à tout document ayant pour ambition de représenter fidèlement le champ concerné, de les prendre en considération. Une telle entreprise nécessite d'être en mesure de saisir l'ampleur de la création normative. Certaines ontologies indiquent tenir compte de telle ou telle interprétation administrative ou jurisprudentielle. C'est le cas de l'ontologie *Consent & Data Management Model*²⁸ (*CC Consent*) laquelle cherche à offrir une représentation des questions de consentement telles qu'elles figurent au sein du RGPD. Les rédacteurs de l'ontologie ont indiqué s'être fondés notamment sur une mise en demeure de la Commission nationale informatique et liberté française sur la validité du consentement. Les auteurs font également une référence générale aux travaux du groupe de travail européen sur la protection des données personnelles mais aussi à des publications de cabinets d'avocats et de différentes autorités nationales²⁹. Le travail de veille semble donc conséquent mais le manque de précision quant à l'usage de ces éléments dans la représentation du droit ne permet ni d'évaluer leurs apports aux représentations, ni de s'assurer de la véracité des informations produites.

Dans ce cadre, il est regrettable que la jurisprudence de la Cour de Justice de l'Union européenne (CJUE) ne soit pas citée en source principale d'interprétation du RGPD. En effet, cette dernière a eu l'occasion de se prononcer sur les méthodes de recueil du consentement au traitement des données personnelles et a rejeté tout consentement présumé, exigeant la preuve d'un comportement actif de la part de la personne concernée³⁰. Les juges s'opposent donc aux cases pré-cochées pour consentir au traitement des données mais aussi aux pratiques contractuelles consistant à induire la personne concernée en erreur en la laissant penser que le contrat ne saurait être conclu sans le consentement au traitement des données personnelles. Les juges ont également invalidé la pratique visant à demander à la personne concernée par le traitement de remplir un formulaire distinct pour s'opposer à l'opération³¹. Par contraste, l'ontologie se borne à ne faire référence qu'au seul comportement actif, reprenant mot pour mot le RGPD : « Le consentement de la personne concernée correspond à toute « toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement »³². Eu égard à l'interprétation de la CJUE, l'entrée de l'ontologie sur le consentement pourrait être enrichie et comprendre également des précisions quant à la qualité de l'information à fournir pour

²⁷ < http://openscience.adaptcentre.ie/projects/GDPRtEXT/dpd_mapping.html >, dernier accès le 20 mars 2021.

²⁸ < <https://openscience.adaptcentre.ie/ontologies/consent/docs/index-en.html#DataSubjectRepresentative> >, dernier accès le 20 mars 2021.

²⁹ Le texte de présentation de l'ontologie précise : « *The primary source of information were the articles and recitals pertaining to consent within the GDPR. Additionally, Article 29 Working Party, which was the official advisory body providing expert opinions regarding data protection, has provided guidelines on consent [1] that assisted in the interpretation of the GDPR. Apart from these, various guidelines and reports published by the Data Protection Offices and legal firms, Handbook on European Data Protection Law, and relevant court laws [2] were also used to understand and formulate technical requirements regarding consent* », para. 3.2, < <http://openscience.adaptcentre.ie/ontologies/gconsent/main.html> > dernier accès le 20 mars 2021.

³⁰ CJUE, décisions du 11 novembre 2020 (C-61/19), décision du 1er octobre 2019 (*Planet49*, aff. C-673/17, ECLI:EU:C:2019:801).

³¹ Considérant 52 de la décision précitée CJUE (n°61/19).

³² Article 4 (11) du RGPD : « consent of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her », voir le texte de l'ontologie : < <http://purl.org/adaptcentre/openscience/ontologies/consent#Consent> >, dernier accès le 20 mars 2021.

recueillir le consentement au traitement des données, à savoir une information « sous une forme compréhensible et aisément accessible ainsi que formulée en des termes clairs et simples, lui permettant [la personne concernée] de déterminer facilement les conséquences de ce consentement, de sorte qu'il soit garanti que celui-ci soit donné en pleine connaissance de cause ». De tels éléments sont centraux pour représenter fidèlement la notion de « consentement » notamment eu égard à la place grandissante des formulaires de consentement en ligne.

La complexité des notions juridiques, mais aussi leurs évolutions, témoignent de la délicate représentation du droit par des ontologies. Dès lors, l'analyse précise des ontologies mais aussi des annotations est un préalable indispensable au déploiement de la justice prédictive laquelle se fonde sur ces outils. Plus encore, d'autres écueils questionnent l'ambition même de la représentation du droit par des ontologies.

II) La représentation du droit : une ambition impossible à atteindre ?

Si la représentation du droit est une tâche complexe, sous certains aspects elle peut s'avérer trop ardue pour être envisagée, posant la question de la pertinence de ces outils, pourtant développés avec succès dans de nombreux autres domaines (y compris en santé). En effet, la volonté de représenter dans un vocabulaire dit naturel et très neutre de grandes notions du droit conduit à les diluer voir à manquer de les représenter avec fidélité. Cet écueil est majeur et les solutions techniques proposées pour y remédier peinent à convaincre de leur efficacité (A). À cette difficulté s'ajoute la délicate traduction des termes, le plus souvent rédigés en anglais, et leur adaptabilité à des systèmes juridiques variés (B).

A) Un vocabulaire inadapté ?

La difficile, voire impossible, représentation du droit par les ontologies est aussi le résultat d'une tension entre les outils sémantiques, d'une part et le droit, d'autre part. En effet, les premiers proposent un vocabulaire généraliste et non spécialisé relevant davantage, dans certains cas, du registre parlé plus que de l'écrit. Bien que permettant de rendre l'annotation flexible et accessible au plus grand nombre, cette stratégie conduit à écarter un vocabulaire spécialisé. Or, si le droit et les savoirs sur le droit reposent sur un langage que l'on pourrait qualifier de « commun », ils prennent toutefois appui sur des définitions spécifiques à l'image de celles des mots « constitution » ou « incapacité ». Ainsi, toute représentation du droit par des ontologies pourra manquer ce degré de finesse alors même que les résultats anticipés des outils d'intelligence artificielle dans le champ de la justice se veulent, eux, précis. Comme le rappelle Philippe Le Tourneau « le Droit est une langue, une langue de précision. La bonne compréhension du droit (et son bon usage) commence par conserver un « sens pur aux mots de la tribu » (pour reprendre une expression de Mallarmé), le juste emploi des mots, pour savoir de quoi l'on parle et éviter les enlacements sémantiques »³³.

Ce niveau de précision semble échapper aux rédacteurs de certaines ontologies. Par exemple, les auteurs de l'ontologie *LKIF* indiquent que « bien que les professionnels du droit utilisent le vocabulaire juridique dans un sens bien plus précis et mesuré que les profanes, il semble que pour la plupart des termes il y ait une compréhension commune suffisante pour les considérer plus ou moins comme similaires, à l'exception d'un certain nombre de termes de base qui ont un sens technique spécifique en droit comme la “ responsabilité ” ou “ le fait légal ” »³⁴. Ici les difficultés sont

³³ Philippe LE TOURNEAU, *Contrats du numérique 2021-2022*, Paris, Dalloz 11^e édition, 2020, chapitre 1, p. 56.

³⁴ « Although legal professionals use the legal vocabulary in a far more precise and careful way than laymen, it appears that for most of these terms there is still a sufficient common understanding to treat them more or less as similar [Lame, 2006], except for a number of basic terms that have a specific legal-technical meaning, such as 'liability' and 'legal fact' », < <http://ceur-ws.org/Vol-321/paper3.pdf> >, dernier accès le 20 mars 2021.

évacuées rapidement en reconnaissant une « compréhension commune » des notions, la technicité de certains termes étant considérée comme une exception à cette généralité exprimée. Or, une telle assimilation entre le sens usuel et le sens juridique est regrettable, les notions étant attachées à des régimes juridiques particuliers qui les éloignent de leur acception en dehors du cadre du droit. C'est le cas du mot « décret » dont le *Vocabulaire Juridique* offre une définition juridique précise. Il s'agit d' « actes administratifs unilatéraux pris par les deux plus hautes autorités de l'État : le Président de la République et le Premier ministre »³⁵. Cette référence aux autorités d'édiction de la norme permet au juriste de saisir à la fois la normativité du document (sa place dans la pyramide des normes) mais aussi d'en comprendre le champ (droit administratif). Le *Dictionnaire Littré* en offre lui une définition usuelle moins précise, le décret étant la « décision par laquelle on ordonne ou règle quelque chose »³⁶. En droit, de nombreux documents unilatéraux entrent dans cette catégorie usuelle, bien qu'ils ne disposent pas tous de la même normativité. En effet, un texte adopté par une association internationale, à l'image de la déclaration d'Helsinki de l'Association Médicale Mondiale relative aux principes éthiques applicables à la recherche médicale³⁷, rédigé avec un vocabulaire directif³⁸ pourra être considéré par erreur comme un décret à l'appui de la définition usuelle. Malgré ce décalage, c'est bien une acception usuelle qui est reprise par l'ontologie *LKIF* pour représenter ce terme, et cela bien maladroitement qui plus est : « le mot décret est souvent utilisé de manière péjorative pour désigner à une décision autoritaire »³⁹.

Pour pallier à cette dilution du sens juridique dans des représentations généralistes, ce n'est pas une seule ontologie qui sera utilisée pour annoter un set de données, mais plusieurs afin d'apporter de la précision. L'on pourra alors retrouver l'ontologie *LKIF*, l'ontologie *ELI* et le vocabulaire *EUROVOC*⁴⁰ associés. Toutefois, ces opérations apparaissent longues, coûteuses et donc complexes à mettre en œuvre⁴¹. Par ailleurs, il faudra évaluer au cas par cas la pertinence de ces ajouts. Une représentation fidèle du droit par les ontologies constitue donc un objectif (trop) complexe à atteindre mais aussi à évaluer et à vérifier.

B) Une traduction illusoire ?

Si la liste des enjeux et écueils à la représentation du droit n'était pas assez longue, il faut encore ajouter la question de la traduction des termes sélectionnés par les ontologies. Tandis que dans d'autres champs les thésaurus et autres indexes sont nombreux, par exemple en médecine où il existe des nomenclatures internationales pour offrir un vocabulaire commun concernant les actes de soin et les traitements⁴², le champ du droit est lui moins poreux à de telles standardisations. En effet, le droit et son ancrage culturel dans l'état où il est émis et interprété peuvent difficilement faire l'objet d'une représentation généraliste utilisant une seule langue, à savoir l'anglais. En pratique, l'annotation imposera implicitement ou explicitement une traduction des termes dans la langue du droit à annoter, traduction qui n'est jamais aisée, des concepts similaires pouvant

³⁵ Entrée « décret », Gérard CORNU (dir.), *Vocabulaire Juridique*, Paris, PUF, 2020.

³⁶ Le nouveau Petit Littré, Paris, éditions Garnier, 2009, p. 527.

³⁷ Association Médicale Mondiale, *Déclaration d'Helsinki relative aux principes éthiques applicables à la recherche médicale impliquant des êtres humains*, 1964.

³⁸ Par exemple, le principe 15 de la Déclaration utilise un vocabulaire prescriptif et n'est pas pour autant un décret : « Une compensation et un traitement adéquats doivent être garantis pour les personnes qui auraient subi un préjudice en raison de leur participation à une recherche ».

³⁹ « The word decree is often used as a derogative term for any authoritarian decision », < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

⁴⁰ Pour une présentation de l'outil voir < <https://op.europa.eu/en/web/eu-vocabularies> >, dernier accès le 20 mars 2021.

⁴¹ Cette conclusion est revenue régulièrement lors de la conférence ENDORSE dont les vidéos sont accessibles en ligne : < <https://op.europa.eu/fr/web/endorse/conference-documents> >, dernier accès le 20 mars 2021.

⁴² Voir les travaux de l'Agence du numérique en santé sur ces terminologies sémantiques : < <https://esante.gouv.fr/interoperabilite/espace-des-terminologies-de-sante> >, dernier accès le 20 mars 2021.

connaître des acceptions différentes selon les ordres juridiques. Partant, les risques d'erreurs dans l'annotation sont démultipliés. L'usage de l'entrée « directive » de l'ontologie *LKIF*, mentionnée plus haut, illustre ces difficultés, le terme faisant référence à des réalités distinctes selon les ordres juridiques. L'entreprise qu'est la traduction juridique est donc périlleuse tant elle est particulière et se heurte à de nombreux obstacles à l'image de l'absence d'équivalence entre les langues mais aussi entre les droits⁴³. Comme l'a démontré Simone Glanert, la traduction peut alors être un casse-tête et cela même pour les mots les plus simples⁴⁴. Aussi, la question de la traductibilité des termes et de leur application dans un ordre juridique particulier met en lumière les obstacles s'opposant à la production d'ontologies à prétention universelle.

Si les traductions posent des difficultés particulières, l'utilisation d'un vocabulaire commun pour décrire les droits nationaux est elle aussi problématique. En effet, cette utilisation peut conduire à opérer une uniformisation des concepts de droit par l'annotation des textes, les représentations considérées comme communes prenant le pas sur les catégories juridiques nationales. Pourtant, les comparatistes ont également montré combien l'uniformisation du droit est une tâche délicate en raison des divergences de cultures juridiques⁴⁵. En ce sens, un contrat en droit français est fondé avant tout sur l'échange des volontés, la définition du contrat étant intimement liée à l'expression de ces dernières. L'ontologie *LKIF* définit, elle, le contrat comme un document légal constitué d'une ou de plusieurs règles juridiques toutes formulées exclusivement par une personne physique ou morale⁴⁶. Cette représentation du terme « contrat » est ici très large et, si elle recoupe la notion française de contrat, elle ne l'épouse pas pour autant. L'emploi de l'ontologie pour annoter des contrats français conduirait donc à des erreurs notables et à uniformiser, par l'application d'une définition unique, un concept juridique qui s'y prête mal.

La représentation du droit par des ontologies est donc un travail complexe fondé sur la sélection des termes et leur définition. Ces éléments conditionnent véritablement les résultats des outils de justice prédictive qui dépendent de l'annotation des données. Partant, il devient crucial d'évaluer ces opérations pour s'assurer de leur fiabilité. Cela requiert d'avoir des informations sur les ontologies utilisées pour annoter les données ainsi que sur les méthodes de construction et de mise à jour des documents, information qui n'a pas encore de caractère obligatoire. À cet égard, les récents projets européens relatifs à l'encadrement de l'IA apportent des éléments de réflexions stimulants. Du côté du Conseil de l'Europe, la Commission européenne pour l'efficacité de la justice envisage la certification des outils algorithmiques dans le champ de la justice⁴⁷. Cette certification conduirait à vérifier que les algorithmes respectent la *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*, laquelle comprend un principe de transparence pour « rendre accessibles et compréhensibles les méthodologies de traitement des données » et pour « autoriser les audits externes »⁴⁸. Ce principe pourrait être judicieusement étendu aux ontologies et cela précisément à des fins de vérification des actions menées sur les données qui alimentent l'outil algorithmique. Dans cette même perspective, la proposition de Règlement de la Commission européenne relative à l'encadrement de l'IA publiée en avril 2021 suggère de garder une trace des opérations effectuées sur les données pour les systèmes posant les

⁴³ Voir sur les difficultés de la traduction : Simone GLANERT, *De la traductibilité du droit*, Paris, Dalloz, 2011.

⁴⁴ Voir dernièrement Simone GLANERT, « "Aujourd'hui, maman est morte" : traduction littéraire et droit comparé », *Revue Droit & Littérature*, vol. 4, no. 1, 2020, pp. 371-389.

⁴⁵ Voir notamment Pierre LEGRAND (dir.), *Comparer les droits, résolument*, Paris, PUF, 2009 et plus particulièrement p. 209 et s.

⁴⁶ < <https://github.com/RinkeHoekstra/lkif-core/blob/master/norm.owl> >, dernier accès le 20 mars 2021.

⁴⁷ < <https://rm.coe.int/draft-agenda-23sept2019-certification-work-2/1680973af5> >, dernier accès le 20 mars 2021.

⁴⁸ Conseil de l'Europe, *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement*, 2019. Voir le principe n°5.

plus grands risques pour les personnes (article 10 (2) (c))⁴⁹. Cette proposition, qui s'appliquera à n'en pas douter aux dispositifs d'IA dans le champ de la justice, peut, suivant son interprétation, conduire à conserver les informations sur les opérations d'annotation des données. Il s'agit donc d'une première pierre, complémentaire à l'approche développée par le Conseil de l'Europe, et nécessaire pour assurer la traçabilité des opérations sémantiques, traçabilité qui devra, pour être véritablement utile, se doubler d'une analyse critique de la formulation des ontologies.

⁴⁹ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 21 avril 2021, COM(2021) 206 final 2021/0106 (COD).