

Chapitre 1

Analyse automatique de corpus judiciaires

L'étude bibliographique de ce chapitre est focalisée sur l'application de techniques d'analyse de données textuelles judiciaires. L'état de l'art plus technique est décrit dans les chapitres qui traitent, dans la suite, des méthodes que nous avons mises en œuvre.

1.1 Introduction

Les deux grands paradigmes de jugement se distinguent par l'importance qu'ils accordent aux règles juridiques (Tumonis, 2012). D'une part, les adeptes du Formalisme Juridique, plus pertinent dans le droit civil, considèrent que toutes les considérations normatives ont été incorporées dans les lois par leurs auteurs. D'autre part, l'école du Réalisme Juridique, plus proche du « *Common Law* », permet un pouvoir discrétionnaire entre les jugements en raisonnant selon le cas. Les premières tentatives d'anticipation des comportements judiciaires s'appuyaient sur une formalisation des lois. Le « droit computationnel » est une sous discipline de l'« informatique juridique¹ » qui en est née. Il s'intéresse, en effet, au raisonnement juridique automatique axé sur la représentation sémantique riche et plus formelle de la loi, des régulations, et modalités de contrat (Love and Genesereth, 2005). Il vise à réduire la taille et la complexité de la loi pour la rendre plus accessible. Plus précisément, le

1. Application des techniques modernes de l'informatique à l'environnement juridique, et par conséquent aux organisations liées au droit

« droit computationnel » propose des systèmes répondant à différentes questions, comme « Quel montant de taxe dois-je payer cette année ? » (planification juridique), « Cette réglementation contient-elle des règles en contradiction » (analyse réglementaire), « L'entreprise respecte-t-elle la loi ? » (vérification de la conformité) (Genesereth, 2015). Les techniques pro Formalisme Juridique étaient déjà critiquées au début des années 60, d'être limitées car se focalisant excessivement sur les règles juridiques qui ne représentent qu'une partie de l'institution juridique (Llewellyn, 1962). Pour analyser le comportement judiciaire, plusieurs variables plus ou moins contrôlables, comme le temps, le lieu et les circonstances, doivent aussi être nécessairement prise en compte (Ulmer, 1963). Etant donné que les juristes s'appuient sur la recherche de précédents, Ulmer (1963) conseille de se concentrer sur les motifs réguliers que comprennent les données pour réaliser des analyses quantitatives. Nous exploitons ainsi la masse de décisions pour identifier de telles régularités car une collection suffisante d'une certaine forme de données révèle des motifs qui une fois observés sont projetables dans le futur (Ulmer, 1963). Il s'agit donc de raisonnements à base de cas qui se distinguent du raisonnement à base de règles.

Les premiers outils automatiques d'anticipation des décisions étaient généralement des systèmes experts juridiques. Ces derniers raisonnent sur de nouvelles affaires en imitant la prise de décision humaine par la logique en général et souvent par analogie. Ils s'appuient sur un raisonnement à base de règle c'est-à-dire à partir d'une représentation formelle des connaissances des experts ou du domaine. En droit, il s'agit de la connaissance qu'à l'expert des normes juridiques et de l'ordre des questions à traiter lors du raisonnement sur un cas (appris par expérience). Le modèle explicite de domaine nécessaire ici se trouve dans une base de connaissances où les normes juridiques sont représentées sous forme de « SI ... ALORS ... », et les faits sont généralement représentés dans la logique de prédicat. Un système expert juridique doit s'appuyer sur une base de connaissances juridiques exhaustive et disposer

d'un moteur d'inférence capable de trouver les règles pertinentes et le moyen efficace, par déduction, **et** les appliquer afin d'obtenir la solution du cas actuel aussi rapidement que possible. Les systèmes experts ont échoué dans leur tentative de prédire les décisions de justice (Leith, 2010). La première raison découle de ce que Berka (2011) a appelé le « goulot d'acquisition de connaissances » c'est-à-dire le problème d'obtention des connaissances spécifiques à un domaine d'expertise sous la forme de règles suffisamment générales. L'autre raison tient à l'interprétation ouverte du droit et à la complexité de la formalisation applicable sans tenir compte des particularités de l'affaire.

Contrairement au raisonnement à base de règle, le raisonnement à base de cas concerne une recherche de solution, une classification ou toute autre inférence pour un cas courant à partir de l'analyse d'anciens cas et de leur solution (Moens, 2002). Un tel système juridique résout les nouveaux cas en rapprochant les cas déjà réglés et en adaptant leurs décisions (Berka, 2011). Le raisonnement fondé sur des cas connaît un succès croissant dans la prédiction de l'issue d'affaire **plus** aux États-Unis qu'ailleurs. Pour exemple, Katz et al. (2014) entraînent des forêts aléatoires sur les cas de 1946-1953 pour prédire si la Cour Suprême des États-Unis infirmera ou confirmera une décision de juridiction inférieure. Ils ont réussi à atteindre 69,7% des décisions finales pour 7 700 cas des années 1953-2013 ; des résultats qu'ils ont légèrement améliorés plus récemment en augmentant le nombre d'arbres et la quantité de données (Katz et al., 2017). Pour le même type de tâche, Walzl et al. (2017a) utilisent des techniques de TALN et extraient automatiquement moins de caractéristiques que (Katz et al., 2014) à partir des décisions d'appel de la Cour Fiscale Allemande (11 contre 244). Ils obtiennent des valeurs de f1-mesures entre 0,53 et 0,58 (validation croisée à 10 itérations) pour la prédiction de la confirmation ou l'infirmerie d'un jugement en appel avec un classifieur bayésien naïf. D'autre part, Ashley and Brüninghaus (2009) ont obtenu une précision de 91,8% en tentant de prédire la partie **(**plaignant/défendeur) qui sera favorisée à l'issue d'affaires d'appropriation illicite de secrets commer-

ciaux. Contrairement à (Katz et al., 2014) qui catégorisent les caractéristiques de valeurs prédéfinies pour caractériser la décision débattue, les tribunaux et les juges (opinions politiques, origine de l'affaire, identifiant du juge, raison et sens du dispositif de la cour inférieure), Ashley and Brüninghaus (2009) identifient, par classification, des facteurs pouvant influencer la décision. Les valeurs des caractéristiques de ces différents travaux sont prédéfinis et très limités, et ne reflètent pas, par conséquent, la grande diversité de catégories qu'on peut retrouver dans les décisions.

Notre objectif est d'alimenter les analyses quantitatives de corpus jurisprudentiels en proposant des méthodes d'extraction de connaissances pertinentes telles que les références des affaires (juge, date, juridiction, etc.), les règles juridiques associées, les demandes de parties, les réponses des tribunaux, et les liens entre ces données. Les juges apportent une réponse à chaque demande, et par conséquent une partie peut voir ses demandes soit toutes acceptées ou rejetées, soit une partiellement accordées. Un juriste sera donc plus intéressé à formuler et défendre les demandes qui ont de meilleures chances d'être acceptées pour un type de contentieux précis plutôt que de prévoir une victoire du procès. C'est la raison pour laquelle notre analyse se situe à un niveau de granularité plus fin (la demande), contrairement aux travaux sur la prédiction qui traitent d'un résultat global sur la décision (par ex. confirmer/infirmier ou gagner/perdre). L'identification de ces diverses connaissances est possible par l'analyse sémantique des textes judiciaires grâce aux méthodes du TALN et de l'analyse (ou fouille) de données textuelles. Cependant, l'application de ces techniques exigent certaines adaptations pour surmonter les divers défis décrits par Nazarenko and Wyner (2017) : textes très longs et en grande quantité, corpus régulièrement mis à jour, influence subjective de facteurs sociaux et d'opinions politiques, couvertures de problématiques économiques, sociales, politiques très variées, langage complexe, etc. Dans la suite, nous passons en revue des travaux qui ont été menés dans ce sens pour traiter de problématiques proches des

nôtres, en particulier celles décrites précédemment dans l'introduction (Section § ii.b).

1.2 Annotation et extraction d'information

L'annotation consiste à enrichir les documents pour les préparer à d'autres analyses, faciliter la recherche d'affaires pertinentes, et faire la lumière sur des connaissances linguistiques sous-jacentes au raisonnement juridique. Les éléments annotés peuvent être de très courts segments de texte mentionnant des entités juridiques (Waltl et al., 2016; Wyner, 2010) comme la date, le lieu (juridiction), les noms de juges, des citations de loi. L'annotation de passages plus longs consiste à identifier des instances de concepts juridiques plus complexes comme les faits (Wyner, 2010; Wyner and Peters, 2010; Shulayeva et al., 2017), les définitions (Waltl et al., 2016, 2017b), des citations de principes juridiques (Shulayeva et al., 2017), ou des arguments (Wyner et al., 2010).

Différentes méthodes ont été expérimentées pour la reconnaissance d'information dans les documents judiciaires. C'est le cas des modèles probabilistes HMM et CRF que nous étudions dans le chapitre 2. Ils peuvent être combinés à d'autres approches dans un système global. Après avoir segmenté les documents à l'aide d'un modèle CRF, Dozier et al. (2010) ont combiné plusieurs approches pour reconnaître des entités dans les décisions de la Cour Suprême des États-Unis. Ils ont définis des détecteurs distincts à base de règles pour identifier séparément la juridiction (zone géographique), le type de document, et les noms des juges, en plus de l'introduction d'une recherche lexicale pour détecter la cour, ainsi qu'un classifieur entraîné pour reconnaître le titre. Ces différents détecteurs ont atteint des performances prometteuses, mais avec des rappels limités entre 72% et 87%. Suivant la complexité des éléments à extraire, un système peut exploiter un lexique pour les motifs simples et non-systématiques (indicateurs de mentions de résultats ou de parties) et

les règles pour des motifs plus complexes et systématiques (par ex. noms de juges, énoncés de décisions) (Waltl et al., 2016, 2017b; Wyner, 2010). Cardellino et al. (2017) ont plutôt utilisé un modèle CRF et les réseaux de neurones sur des jugements de la Cour Européenne des Droits de l’Homme. Les basses performances, qu’ils rapportent, illustrent bien la difficulté de la détection d’entités juridiques. Plus récemment encore, Andrew and Tannier (2018) obtiennent de bons résultats en combinant l’extraction d’entités non-juridiques par CRF à celle des relations entre ces dernières par une grammaire GATE JAPE (Thakker et al., 2009) sur des décisions du Luxembourg rédigées en français.

Pour la détection des arguments, par contre, Moens et al. (2007) proposent une classification binaire des phrases : *argumentative* / *non argumentative*. Ils comparent notamment le classifieur bayésien multinomial et le classifieur d’entropie maximum tout en explorant plusieurs caractéristiques textuelles. Mochales and Moens (2008) proposent, pour la même tâche, une approche d’extraction basée sur une formalisation de la structure des arguments dans les jugements par une grammaire sans contexte.

1.3 Classification des jugement

La classification permet d’organiser un corpus en rangeant les documents dans des catégories prédéfinies généralement par des experts. Aletras et al. (2016) identifient s’il y a eu une violation d’un article donné de la convention des droits de l’homme sur les jugements² de la Court Européenne des Droits de l’Hommes (ECHR). Avec un SVM (Machine à Vecteurs de Support) et une représentation vectorielle basée sur les plus fréquents n-grammes et le cluster de leur vecteur de plongement sémantique (word2vec), ils obtiennent une précision moyenne de 79% sur les 3 articles qu’ils ont manipulés. Notons tout de même la sélection particulière des régions du documents où sont

2. HUDOC ECHR Database : <http://hudoc.echr.coe.int>

extraits les n-grammes (circonstances, faits, lois, ...). Cette sélection est un ajustement de la représentation des textes qui paraît nécessaire pour obtenir de bons résultats. La structuration préalable des documents est utile pour réduire le bruit qui occupe généralement plus d'espace que les passages ou éléments d'intérêt. Medvedeva et al. (2018) étendent ces travaux à neuf articles tout démontrant empiriquement, entre autres, la possibilité de prédire la violation des articles sur des périodes futures à celles des données d'entraînement. Şulea et al. (2017a) traitent, d'autre part, l'identification des résultats dans des arrêts³ de la Cour Française de Cassation. Après un essai avec un SVM (Şulea et al., 2017b), ils améliorent les résultats à l'aide d'un classifieur ensembliste de SVM à moyenne de probabilités, parvenant ainsi à des fl-mesures de plus de 95%. Par ailleurs, Ashley and Brüninghaus (2009) entraînent un classifieur (les plus-proches-voisins) pour chacun des 27 facteurs prédéfinis pour savoir s'il s'applique à la décision (phase SMILE). La partie remportant le procès est prédite par un algorithme séquentiel qui compare les parties (plaignant et défendeur) suivant le niveau de préférence des questions juridiques dégagées par les facteurs tel qu'observé dans la base d'entraînement (phase IBP). D'autres catégorisations, comme la formation judiciaire ou la période du prononcé (Şulea et al., 2017b,a), sont toutes aussi utiles pour faciliter la recherche d'information. La classification peut aussi être utilisée à des fins d'évaluation sur d'autres problématiques comme la similarité (Ma et al., 2018).

1.4 Similarité entre décisions judiciaires

La similarité entre textes est indispensable pour des applications qui nécessitent de rapprocher quantitativement des textes traitant de sujets similaires, et resp. éloigner ceux dont les sujets sont différents. La mesure de similarité doit être définie de sorte à rapprocher ou éloigner les documents

3. Documents de <https://www.legifrance.gouv.fr>

suivant l'aspect sémantique qu'on veut révéler. Nair and Wagh (2018) arrivent à exploiter les citations de lois et précédents, car les jugements du « Common Law » citent des décisions d'affaires similaires antérieures. Ils analysent le réseau de citations d'un corpus de 597 documents, à l'aide de règles d'association générées par l'algorithme Apriori pour regrouper les jugements susceptibles d'être cités ensemble. Ils démontrent au travers de scénarios (pas d'évaluation statistique) que les documents qui sont cités fréquemment ensemble sont similaires, et cette relation permet par une transitivité de retrouver les documents pertinents dans une base de données. Certaines métriques traditionnelles, comme la distance cosinus (Thenmozhi et al., 2017), ont été utilisées sur les décisions judiciaires mais pas toujours avec succès. La raison peut venir notamment de la représentation des textes qui doit accentuer l'aspect sous-jacent de la similarité. Ma et al. (2018) proposent donc d'aligner le document sur une ontologie des concepts et relations du corpus judiciaire. L'idée est de calculer la similarité sur un résumé du texte qui compacte le texte uniquement sur les aspects pertinents. Cette méthode permet ainsi de mieux capter la sémantique pure des jugements, d'avoir une meilleure précision, et réduire la complexité temporelle inhérente à l'exploitation de long document notamment lors de l'utilisation de la « distance du déménageur de mot » ou WMD (*Word Mover's Distance*) de Kusner et al. (2015). L'amélioration a été observée sur une tâche de classification avec des jugements Chinois relatifs aux crimes de la circulation routière dans quatre catégories correspondantes à des sentences d'emprisonnement (précision de 90.3% et 92.3% pour le résumé contre 84.8% et 82.4% resp. pour le document original).

Toujours dans l'objectif d'une représentation pertinente des textes, Kumar et al. (2011) proposent quatre méthodes propres aux décisions judiciaires pour l'estimation de la similarité entre deux jugements x et y de la Cour Suprême d'Inde : (i) le cosinus de similarité entre les représentations TF-IDF de x et y (*term frequency - inverse document frequency*) dont tous les termes

présents dans les jugements sont les dimensions (*all-term cosine similarity*), (ii) le cosinus de similarité des termes juridiques réduit les dimensions précédentes uniquement qu'aux termes apparaissant dans un dictionnaire juridique (*legal-term cosine similarity*), (iii) la similarité de couplage bibliographique égal au nombre de citations de jugements qu'ont en commun x et y (*bibliographic coupling similarity*), (iv) la similarité de co-citation qui est le nombre de citations de x et y dans un même jugement (*co-citation similarity*). La similarité étudiée ici est basée sur trois critères : la similarité sur la question discutée, la similarité sur les faits sous-jacents, et l'utilité du document pour les avocats cherchant des documents similaires à une décisions données. Malgré qu'ils aient interprété les résultats sur de très faibles proportions des données utilisées (5/2430 et 18/2430), il en ressort que le cosinus de similarité avec les termes juridiques et le couplage bibliographique correspondent aux valeurs de similarité des experts, contrairement à la similarité basée sur tous les termes du corpus ou sur la co-citation. Thenmozhi et al. (2017) compare aussi la similarité cosinus sur trois représentations différentes des affaires dans le cadre de la campagne de recherche d'affaires antérieures pertinentes IRLed@FIRE2017 : (i) TF-IDF des concepts (noms), (ii) TF-IDF des concepts et relations (verbes), (iii) et la moyenne des *Word2Vec* (Le and Mikolov, 2014) des concepts et relations. Vue les performances qu'ils ont obtenues (précision@10 et rappel@10), la première représentation capte mieux la similarité par rapport à l'utilisation des verbes et de la représentation distribuée.

En bref, la similarité entre documents est utilisée pour répondre à plusieurs tâches, comme par exemple, la recherche des décisions similaires (Thenmozhi et al., 2017), le regroupement non-supervisé des jugements (Raghuveer, 2012) et la classification supervisée de ces derniers (Ma et al., 2018). Ces diverses applications définissent aussi la sémantique juridique liée à la notion de similarité. Parmi les questions liées à la conception d'une mesure de la similarité entre documents, on distingue : la sémantique experte qui fonde

cette similarité, sa métrique de mesure, la représentation des documents, le contexte d'exploitation et les métriques d'évaluation. Les diverses études menées sur la similarité démontrent l'importance de l'abstraction des textes par les concepts soit via l'alignement du document avec une ontologie, soit via la sélection de termes clés.

1.5 Conclusion

En résumé, les travaux portant sur l'analyse automatique des décisions ont donné des résultats encourageant grâce aux éléments spécifiques aux affaires et généralement contenus dans les documents correspondants. Ces éléments peuvent être extraits des décisions grâce aux techniques de TALN et de fouille de texte. L'analyse des données textuelles juridiques a pour but la structuration des documents, l'extraction d'information, et l'organisation sémantique de corpus. Le domaine est très actif depuis déjà plusieurs décennies, au point où des librairies de développement, spécifiques au domaine, commencent à voir le jour (Bommarito et al., 2018). La revue littéraire fait remarquer que le concepteur investit un minimum d'ingénierie d'adaptation que ce soit pour la définition des caractéristiques pertinentes pour les modèles à apprentissage automatique, soit pour définir les règles pour les méthodes à base de règles ou à base de grammaire. Notons aussi l'effort d'évaluation quantitative avec la participation d'experts pour l'annotation d'exemples de référence même pour des tâches qui peuvent paraître subjectives comme la mesure de similarité.