



HAL
open science

Éléments d'état de l'art sur l'extraction et la modélisation de règles formelles à partir de textes légaux

Nathalie Aussenac-Gilles, Jérémy Bouché-Pillon, Pascale Zaraté, Yannick Chevalier, Pierre-Yves Gicquel

► To cite this version:

Nathalie Aussenac-Gilles, Jérémy Bouché-Pillon, Pascale Zaraté, Yannick Chevalier, Pierre-Yves Gicquel. Éléments d'état de l'art sur l'extraction et la modélisation de règles formelles à partir de textes légaux. 34es Journées francophones d'Ingénierie des Connaissances (IC 2023) @ PFIA 2023, Collège SIC (Science de l'Ingénierie des Connaissances) de l'AFIA, Jul 2023, Strasbourg, France. pp.79-84. hal-04153047

HAL Id: hal-04153047

<https://hal.science/hal-04153047>

Submitted on 6 Jul 2023

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.

Éléments d'état de l'art sur l'extraction et la modélisation de règles formelles à partir de textes légaux

J. Bouché-Pillon¹, N. Aussenac-Gilles^{1,2}, P. Zaraté^{1,3}, Y. Chevalier^{1,4}, P-Y. Gicquel^{1,2}

¹ Université de Toulouse - IRIT, France

² CNRS

³ Université Toulouse 1 Capitole

⁴ Université Toulouse 3 Paul Sabatier

jeremy.bouche-pillon@irit.fr

Résumé

L'extraction d'énoncés formels interprétables par une machine à partir de textes en langage naturel est un champ de recherche largement étudié aujourd'hui. Un des cadres d'application de ce domaine de recherche consiste à extraire des règles formelles à partir de textes de lois et de réglementations. Cet article présente une synthèse de différentes approches d'extraction de règles à partir de réglementations légales, puis de représentation formelle de ces règles. A partir de cette synthèse, nous présentons l'ébauche d'un outil d'aide à la décision de partage d'informations sensibles entre différentes organisations, dans l'optique de valider que ces partages soient conformes aux réglementations applicables (RGPD ou directives européennes sur l'usage de l'IA).

Mots-clés

Traitement Automatique des Langues, Extraction de connaissances, Formalisme légal, Logique déontique, Représentation des connaissances, État de l'art.

Abstract

The extraction of machine-interpretable formal statements from natural language texts is a challenging and widely studied research field today. One of the application frameworks of this research field consists in extracting formal rules from legal texts and regulations. This paper presents a synthesis of different approaches to extracting rules from legal regulations, and then to formally representing these rules. From this survey, we present the draft of a tool to help of a decision support tool for sharing sensitive information between different organizations, with the aim of validating that these shares comply with the applicable regulations (GDPR or European directives on the use of AI).

Keywords

Natural Language Processing, Knowledge extraction, Legal formalism, Deontic Logic, Knowledge representation, Survey.

1 Introduction

La dénotation de la sémantique d'un texte en langue naturelle dans une langue formelle est un domaine difficile mais aussi très étudié. Avoir une extraction exacte est d'une importance primordiale en particulier lorsque l'on souhaite extraire les règles opérationnelles des lois et règlements. L'extraction de représentations formelles et en particulier de règles à partir de textes juridiques peut permettre entre autres de réguler l'accès à l'information. Cependant, la formalisation des textes de lois et réglementations demeure problématique alors qu'elle présente un enjeu majeur pour les professionnels du droit mais aussi pour les citoyens. Durant les dernière décennies, ce problème a connu de nombreuses avancées avec le développement de nouvelles approches [13, 14].

Nous nous intéressons plus particulièrement au droit des données, relatif à leur partage, leur protection ou leur traitement algorithmique par apprentissage automatique. En effet, avec l'augmentation significative des quantités et des types de données acquises, stockées, traitées et échangées par des entreprises ou organisations au cours des dernières décennies, un fort besoin de légifération a émergé afin d'encadrer leur utilisation de ces données. Plusieurs lois et réglementations ont déjà été mises en application, comme par exemple le "Règlement Général sur la Protection des Données" (RGPD)¹ à l'échelle européenne, et de nombreux autres textes seront amenés à être votés à l'avenir, comme l'"Artificial Intelligence Act" (IA Act). Toute organisation traitant des données doit se conformer et s'adapter aux nouvelles lois au fur et à mesure de leur adoption par un gouvernement, par exemple en faisant appel à un expert juridique. S'assurer que chaque traitement de données réalisé est conforme aux réglementations est une tâche laborieuse et répétitive qui nécessite des ajustements à chaque entrée en vigueur d'un nouveau texte. Le recours à des juristes s'impose alors, mais cette démarche peut ralentir le partage dans des situations où la rapidité est parfois critique. Disposer

1. General Data Protection Regulation (GDPR) – Official Legal Text, <https://gdpr-info.eu/>

d'un outil permettant d'automatiser l'extraction et la formalisation d'énoncés légaux en vue de les utiliser dans un système d'aide permettrait de faciliter le suivi du respect des réglementations.

Une application concrète que nous envisageons ici pour répondre à ce besoin est la conception d'un système d'aide à la décision pour des partages de données entre autorités chargées de faire respecter la loi. Ces autorités sont naturellement soumises à de nombreuses réglementations, en l'occurrence des réglementations portant sur les conditions dans lesquelles elles peuvent acquérir, traiter et partager des informations entre elles (par exemple pour obtenir des preuves en possession d'une autre autorité dans le cadre d'une enquête judiciaire). Or, la diversité de situations de partage d'informations et le flou inhérent à l'interprétation des textes légaux rendent particulièrement difficile le codage des obligations, permissions et interdictions par des informaticiens. Nous proposons ainsi une approche en 2 étapes. Dans la première, lorsqu'un texte ou une jurisprudence est identifié applicable sans ambiguïté, une décision en accord avec ce texte est prise. Dans la seconde étape, en cas d'ambiguïté, le contexte légal du partage est fourni à l'outil pour suggérer une décision qui soit en accord avec des décisions prises par le passé dans des situations similaires. Ainsi, afin de prendre en compte à la fois les textes légaux applicables et l'historique des partages, il est important de disposer d'un outil automatisant la traduction de ces textes juridiques écrits en langage naturel en règles opérationnelles utilisables par un programme de partage.

Le problème de formalisation mis en évidence précédemment se décompose en 2 étapes complémentaires, illustrées sur la figure 1 : (i) Extraire et identifier les différents éléments qui constituent les règles à partir de textes légaux. On traite ici de questionnements relatifs au domaine du Traitement Automatique des Langues. (ii) Adopter un formalisme dans lequel représenter ces règles dans l'optique de les utiliser dans un outil capable de vérifier la conformité d'un traitement de données aux réglementations en vigueur. Cette facette se rattache au domaine de la Représentation des connaissances.

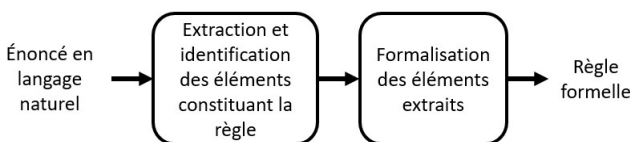


FIGURE 1 – Chaîne de traitement étudiée

A ces deux étapes, s'ajoute une phase préliminaire de caractérisation des différents éléments à extraire des textes. En effet, cette phase est indispensable pour mener une recherche précise, efficace et pertinente.

Les états de l'art récents portant sur les 3 aspects ainsi étudiés sont rares, mais on peut cependant trouver quelques états de l'art sur différents formalismes, comme [11] par exemple.

Nous structurons donc la suite de l'article en trois parties

dressant un état de l'art sur la nature des éléments formant une règle juridique (2), sur les approches permettant de les extraire de textes (3) et enfin sur les formalismes définis ou utilisés pour les représenter (4). Nous esquissons en conclusion les caractéristiques d'un outil d'aide à la décision basé sur ce type de règles formalisées.

2 Les règles légales et la déontique

Afin de complètement contextualiser et caractériser les règles que nous cherchons à formaliser, nous commençons par définir ce que sont la logique et les modalités déontiques qui interviennent dans les règles légales avant de lister les différents éléments qui constituent une règle légale.

2.1 Logique et modalités déontiques

La *déontique* est une étude systématique des propriétés formelles vérifiées par des notions juridiques, comme par exemple celles de droit et d'obligation. La *logique déontique*, elle, tente ainsi de formaliser les rapports qui existent entre les caractéristiques d'une loi, les *modalités déontiques*. Il s'avère que le dénombrement de ces modalités peut être sujet à débat. En effet, dans une vision logicienne répandue héritée d'Aristote, les modalités déontiques sont au nombre de quatre, organisées en deux couples : l'interdiction et son contraire, la permission (droit de faire), d'une part, et l'obligation et son contraire, le facultatif (droit de ne pas faire), d'autre part [15]. [15] souligne cependant qu'en pratique, le "droit de faire" et le "droit de ne pas faire" se recourent et se superposent en une unique notion à la fois négation de l'obligation et de l'interdiction : une nouvelle interprétation de "permission" qui désigne ce qui n'est ni obligatoire, ni interdit.

On considère cependant pour notre étude que la formulation des textes de loi nous permet d'effectivement distinguer "permission" et "facultatif" : on choisit de classer les formulations du style "S peut effectuer l'action A" comme une permission et "S peut ne pas effectuer l'action A" comme un facultatif.

2.2 Règles légales : composants

Nous dressons ici un bilan des différents éléments ou concepts susceptibles de se trouver dans une règle légale. Dans la littérature, les éléments généralement recherchés dans les textes en vue de la formalisation de règles légales sont traités à différents niveaux de granularité. En effet, certaines approches réalisent une analyse grammaticale et sémantique très fine dans laquelle chaque groupe de mots a un rôle et un sens précis avec des liens logiques reliant ces groupes [2, 3, 4, 7]. D'autres approches, plus générales, limitent l'analyse à l'identification de liens logiques entre groupes de mots [6]. D'autres encore se contentent d'identifier les modalités déontiques dans l'optique de classifier les règles légales [1]. Il s'avère aussi que les approches réalisant une analyse poussée identifient des éléments sensiblement différents mais en réalité complémentaires.

Ainsi, une analyse fine du contenu des règles légales permet

de repérer des éléments systématiquement présents, comme ceux mentionnés dans [3] :

- **l'"agent" de la règle** : Il s'agit d'un rôle sémantique rattaché au rôle grammatical de "sujet" dans une phrase. L'"agent" est ainsi l'entité qui est soumis à la règle.

- **le "thème" de la règle** : Analogiquement à l'"agent", le "thème" est un rôle sémantique rattaché au rôle grammatical d'"objet" dans une phrase. Le "thème" est donc l'élément sur lequel s'applique la règle.

- **les termes exprimant l'une des 4 modalités déontiques** : Bien que la logique déontique repose sur les 4 modalités présentées précédemment, il est assez rare que les approches repérées dans l'état de l'art s'attachent à identifier les 4 modalités. Le plus souvent, elles se focalisent sur seulement 2 ou 3 d'entre elles. Ainsi, toutes les approches étudiées ici traitent soit les modalités d'"obligation" et de "permission" [1, 3, 6], soit ces 2 modalités avec l'"interdiction" en plus [2, 7], mais aucune ne s'attarde sur la distinction entre permission et facultatif.

D'une part, le fait que les "interdictions" ne soient pas traitées dans toutes les études s'explique par le fait que beaucoup de textes de lois ne contiennent effectivement pas d'"interdictions". En effet, certains textes n'"interdisent" souvent rien à proprement parler mais définissent des sanctions à certaines actions. D'autre part, lorsque des approches traitent de "permission", il s'agit de la "permission" telle que vue dans [15] et qui désigne donc ce qui n'est ni obligatoire ni interdit.

Il existe de nombreux termes permettant d'exprimer soit une obligation, soit une interdiction, soit une permission, soit un facultatif. Ces termes sont principalement mais non exclusivement des verbes comme "devoir" (obligation), "pouvoir" (facultatif ou permission selon la présence ou non d'une négation dans le "thème") ou "être interdit de" (interdiction).

- **les verbes principaux** : Il s'agit des verbes qui peuvent se trouver dans les textes étudiés mais qui ne sont pas rattachés à l'une des 4 modalités déontiques. Ils permettent de préciser des actions qui doivent être réalisées en conséquence de l'application de la règle ou de contextualiser la règle.

Enfin, d'autres éléments peuvent être présents dans certaines règles déontiques, tels que ceux identifiés dans [3] et dans [7] :

- **des clauses d'exception** : Un élément essentiel dans l'interprétation de règles déontiques est la présence de clauses d'exceptions qui spécifient des conditions particulières dans lesquelles la règle ne s'applique pas, ou s'applique différemment. L'identification de ces clauses d'exception, bien que cruciale, s'avère être une des difficultés principales rencontrées dans l'analyse de règles déontiques, notamment parce qu'elles peuvent se présenter sous la forme d'énumérations. Cet aspect est particulièrement développé dans [3].

- **des phrases conditionnelles** : Les phrases conditionnelles permettent d'exprimer des structures logiques de

type "Si A alors B". Ce genre de phrase est donc caractérisé par deux types d'éléments : les antécédents ("A") et les conséquents ("B"). Comme pour les clauses d'exception, une des difficultés rencontrées lors de l'identification des structures conditionnelles dans les règles déontiques vient du fait que les antécédents peuvent se trouver sous la forme d'énumérations ou de listes [3, 6, 11]. Bien que ces études ne considèrent les listes ou énumérations que dans les antécédents, il ne faudrait pas exclure la possibilité de rencontrer aussi des structures conditionnelles avec des conséquents sous forme d'énumérations.

- **les références croisées** : [6] mentionne la présence dans de nombreux paragraphes de textes de lois de références à d'autres paragraphes du même article, à d'autres articles du même texte, voire à d'autres textes dans la législation.

4. Lorsque les objets, documents ou données concernés sont déjà pertinents pour d'autres procédures, l'autorité d'exécution peut, à la demande expresse de l'autorité d'émission et après consultation de celle-ci, transférer temporairement ces éléments de preuve, à condition qu'ils soient renvoyés à l'État d'exécution dès qu'ils ne sont plus nécessaires à l'État d'émission ou à tout autre moment ou toute autre occasion convenus entre les autorités compétentes.

FIGURE 2 – Paragraphe 4 de l'article 13 de la directive 2014/41/UE

Par exemple, en analysant le paragraphe 4 de l'article 13 de la "DIRECTIVE 2014/41/UE DU PARLEMENT EUROPÉEN ET DU CONSEIL concernant la décision d'enquête européenne en matière pénale", illustré sur la figure 2, on identifie les éléments suivants :

- **"L'autorité d'exécution" est l'"agent"**. En effet, c'est elle qui doit appliquer la règle.

- **"transférer temporairement ces éléments de preuve" est le "thème"**. C'est ce sur quoi porte la règle. Entre autres, on peut identifier "transférer" comme un des "verbes principaux".

- **"peut" est le verbe exprimant la modalité déontique**. Il s'agit ici d'une permission.

- **La règle s'inscrit ici dans une structure conditionnelle sous la forme "Lorsque A, B"**. On a donc comme antécédent "les objets ou données concernées sont déjà pertinents pour d'autres procédures, et comme conséquent "l'autorité d'exécution peut transférer temporairement ces éléments de preuve".

- **Notons que d'autres propositions subordonnées permettent de contextualiser d'avantage la règle en rajoutant des conditions qui viennent ici étoffer la partie "antécédent" de la structure conditionnelle** : "à la demande expresse de l'autorité d'émission et après consultation de celle-ci" et "à condition qu'ils soient renvoyés à l'État d'exécution dès qu'ils ne sont plus nécessaires à l'État d'émission ou à tout autre moment ou toute autre occasion convenus entre les autorités compétentes". On peut également remarquer dans cet exemple que l'exécution de l'action permise crée une autre obligation pour

l'état d'exécution, et que cette obligation n'est pas introduite par un verbe spécifique mais par des formulations "à condition que" et "dès que", ce qui rend l'analyse complexe.

Notons que l'exemple traité ici est en français mais que la plupart des approches étudiées ont été conçues pour analyser des textes législatifs en anglais. Par conséquent les outils utilisés dans ces approches, que ce soient les ontologies, les analyseurs syntaxiques, lexicaux ou grammaticaux, sont dédiés à l'analyse de la langue anglaise. Ce papier ayant pour but d'explorer les approches existantes et non les outils utilisés dans ces approches, toutes les approches mentionnées sont donc réalisables pour des textes dans n'importe quelle langue à condition de disposer d'outils permettant des opérations analogues dans cette langue. Les performances obtenues peuvent cependant varier étant donné que la performance globale dépend des performances individuelles des outils intervenant dans l'analyse.

3 Extraction des éléments identifiés

Une fois caractérisés, les éléments constituant les règles déontiques doivent être extraits des textes en langage naturel. Dans un premier temps, si le texte à analyser n'est pas directement accessible au format texte, mais par exemple en PDF, il est bien sûr nécessaire d'utiliser un outil de conversion, comme c'est le cas dans [2] ou [7]. Une fois le texte disponible au bon format, plusieurs types d'analyses sont possibles.

Approches à base de modèles de langue Par exemple, différents modèles dérivés de BERT sont utilisés dans [1] afin de classer les différents énoncés d'un texte selon, entre autres, les modalités de permission et d'obligation. En amont de cette classification par modèle de langage, [1] réalise une annotation des textes avec les ontologies AkomaNtoso² et LegalRuleML [8] permettant de représenter respectivement la structure de textes légaux et les concepts présents dans les textes légaux. Cette approche, bien que ne permettant pas une extraction de tous les éléments essentiels mentionnés dans la section 2, offre des bons résultats de classification pour identifier quelle modalité déontique est utilisée. L'obligation et la permission étant les seules modalités traitées dans cette approche, il pourrait être intéressant de l'étendre à la modalité d'interdiction qui intervient dans de nombreux textes de loi.

Approches basées sur des chaînes de traitement plus complètes. Ainsi [3] propose une approche purement linguistique (par opposition aux approches utilisant de l'apprentissage). Après de nombreux pré-traitements lors desquels le texte analysé est divisé en sections exploitables, un grand nombre d'opérations sont réalisées via la plateforme GATE [9] qui offre pléthore d'opérations : tokenisation, fractionnement de phrases, étiquetage morpho-syntaxique, distinction des groupes verbaux et

nominaux, utilisation d'un parser. L'analyse utilise le Stanford Parser³ qui est une référence en terme d'analyse de la langue anglaise. Elle s'appuie aussi sur des nomenclatures pour identifier les termes se rattachant aux différentes modalités déontiques ainsi que les termes marquant des structures d'exceptions ou des structures conditionnelles.

Approches reposant sur une analyse sémantique des textes à l'aide d'ontologies de domaines. C'est notamment le cas dans [4] qui utilise dans un premier temps un "modèle documentaire" pour identifier quels morceaux de texte pourront être sémantiquement annotés, chaque morceau de texte étant une "unité documentaire", puis dans un second temps un "modèle sémantique" pour associer une "unité sémantique", qui peut être soit une entité d'une ontologie soit une règle sémantique, à chaque "unité documentaire" identifiée précédemment.

Approches combinant les travaux précédent, comme l'analyse à deux niveaux réalisée dans [6] :

Au niveau du document analysé dans sa globalité La structure hiérarchique du texte est identifiée et le texte se trouve ainsi découpé en sections, sous-sections, articles et phrases. En outre, certains termes spécifiques sont extraits avec leur définition grâce à une analyse lexicale. Ces termes et leur définition sont ensuite associés dans un lexique de référence qui sera réutilisé lors de la génération des règles formelles dans un souci de désambiguïsation.

Au niveau des phrases prises séparément Dans un second temps, chaque phrase peut être découpée en clauses pour réduire sa complexité puis, à partir des informations issues du niveau d'analyse précédent, un module de "Meaning Representation" permet d'identifier des prédicats logiques et leurs arguments pour créer une représentation formelle simple intermédiaire avant le formalisme final. Enfin la génération des règles dans le formalisme final s'appuie sur une ontologie regroupant les concepts qui renvoient aux modalités déontiques de permission et d'obligation. Comme mentionné pour [1], il pourrait être intéressant d'étendre cette analyse à la modalité d'interdiction.

L'approche décrite dans [7] combine également des aspects linguistiques et sémantiques dans l'analyse du texte. En effet, après un découpage du texte en phrases, une analyse sémantique permet d'abord de générer une représentation arborescente de la structure grammaticale des phrases à travers l'utilisation d'"Universal Dependencies". Cet arbre est ensuite traduit en un graphe sémantique via le framework "4lang"⁴. Enfin, de cette représentation en graphes sémantiques, une grammaire IRTG ("Interpreted Regular Tree Grammars") est utilisée pour extraire les éléments nécessaires à l'écriture de règles qui sont ensuite associés les uns avec les autres via une heuristique simple pour obtenir des règles formelles.

3. Software > Stanford Parser, <https://nlp.stanford.edu/software/lex-parser.shtml>

4. 4lang concept lexicon, <http://hlt.sztaki.hu/resources/4lang/>

2. Akoma Ntoso | Akoma Ntoso Site, <http://www.akomantoso.org>

Enfin, [2] propose la chaîne de traitement la plus complète, menant 2 analyses en parallèle et combinant leurs résultats à la fin. Une première étape commune aux 2 analyses consiste à reconstituer des phrases entières en exploitant la structure arborescente des textes de loi. En effet, les phrases ont tendance à se retrouver scindées en différentes propositions qui peuvent se retrouver dans une énumération. Cette reconstitution des phrases s'appuie sur une ontologie basique décrivant la structure arborescente des phrases. Après cette étape, la chaîne de traitement se scinde en 2 branches :

- *Une approche logique* applique d'abord un parser dit "CCG" pour "Combinatory Categorical Grammar". Ce parser réalise une analyse logique de chaque phrase et génère des représentations sémantiques appelées "Discourse Representation Structures" (DRS). Les DRS s'apparentent à des formules logiques du 1er ordre et permettent de trouver certaines connections sémantiques entre les mots d'une phrase. Ces représentations sont ensuite exploitées par le framework "Boxer"⁵ pour extraire les relations logiques qui relient différents blocs de texte. Des règles formelles sont alors générées à partir de ces éléments.

- *Une approche syntaxique* s'en suit, qui commence par appliquer le Stanford Parser aux phrases reconstituées pour en obtenir une représentation grammaticale. Puis les phrases sont découpées en "termes", des groupes de mots qui correspondent en général aux différentes propositions principales et relatives qui constituent chaque phrase. Chacun de ces "termes" est ensuite annoté avec des marqueurs déontiques indiquant s'il exprime une permission, une obligation, une interdiction, ou n'exprime pas de modalité déontique. Enfin, les différents "termes" sont combinés par correspondance à des structures logiques pré-définies pour générer les règles formelles.

Une dernière étape consiste à rassembler et comparer les règles issues de ces 2 branches pour obtenir un unique ensemble de règles.

4 Représentation formelle des règles

Le dernier aspect qui doit être étudié est le formalisme dans lequel représenter les règles qui ont été extraites des textes. Le formalisme choisi dépend bien sûr en grande partie de l'application envisagée pour les règles formelles générées.

Selon [12] 3 critères ont été déterminés : L'expressivité, la réutilisabilité et la compréhension des règles par l'utilisateur, afin de trouver parmi les modèles suivants : ABAC, OBAC, RBAC, OrBAC, Multi-OrBAC, KaOS, le meilleur pour modéliser des règles juridiques européennes sur le partage d'informations en matière de criminalité.

Ensuite, certaines des approches mentionnées plus tôt utilisent des représentations inspirées ou dérivées de logiques du premier ordre ou de formules en lambda calcul pour formaliser les règles. C'est le cas dans [11] où, après un état de l'art de différents langages et formats et de leur

conformité à un ensemble de critères, le langage "Legal Knowledge Interchange Format" (LKIF) est présenté. Il se base sur des axiomes qui sont des formules en logique du premier ordre et sur des règles d'inférences. L'article mentionne dans son état de l'art "Rule Markup Language" (RuleML), "Semantics of Business Vocabulary and Business Rules" (SBVR), "Semantic Web Rule Language" (SWRL) et "Rule Interchange Format" (RIF) et souligne les défauts et manquements de chacun de ces langages.

On trouve aussi dans [7] une formalisation sous forme de logique déontique dyadique, qui est une logique dédiée à la représentation de structures de la forme *Il est obligatoire de faire A étant donné B* qui peut se représenter ainsi : $O(A,B)$. [7] choisit ce formalisme car il s'avère très adapté au cas particulier du texte qui y est formalisé. Cependant, cette représentation s'avère moins adaptée à des textes complexes et très descriptifs. Il est ensuite possible de développer un prouveur à partir de cette logique pour vérifier ce que l'on souhaite.

D'autres approches utilisent d'autres formes de formalisme existants comme par exemple dans [3] où le formalisme utilisé n'est pas exactement un formalisme de représentation des règles mais un formalisme d'annotation du texte. En effet, ce sont des règles JAPE ("Java Annotation Patterns Engine")⁶ qui sont utilisées dans la plateforme GATE [9] pour annoter le texte de façon à identifier les différents éléments constituant une règle déontique. Bien qu'il ne soit pas ici directement question d'une représentation formelle des règles déontiques dans leur totalité, il est facile d'imaginer regrouper les annotations obtenues en une représentation formelle adaptée.

C'est également le cas de [6] qui utilise d'abord une représentation en lambda calcul comme représentation intermédiaire puis exprime ces règles en un formalisme plus lisible, ici du PCL [10], un langage conçu pour la modélisation de processus métier en intégrant des contraintes normatives. Comme avec la représentation dans [7], ce formalisme offre la possibilité de mener des raisonnements pour vérifier le respect des règles.

Il est enfin possible de définir son propre formalisme, comme c'est le cas dans [2] où les règles formelles se représentent sous la forme :

Terme1 => [Permission] NON Terme2 Ce formalisme, bien que très simple à comprendre ne semble pas être dérivé d'une logique existante. De fait, mener des raisonnements à partir de ce formalisme n'est pas trivial et nécessiterait un travail supplémentaire que nous ne détaillerons pas ici.

5 Conclusion

Nous avons présenté un état de l'art d'approches d'extraction et de modélisation de règles formelles à partir de textes de lois écrits en langage naturel. Une analyse du pro-

5. jgordon/boxer : C&C parser & Boxer, <https://github.com/jgordon/boxer>

6. JAPE : Regular Expressions over Annotations, <https://gate.ac.uk/sale/tao/splitch8.html>

blème nous a permis de le structurer selon 3 aspects essentiels : l'identification des éléments constituant des règles déontiques, l'extraction de ces éléments et des liens qui les unissent et le choix d'une représentation formelle des règles. La plupart des approches présentées sont en accord sur les différents éléments qui peuvent constituer une règle déontique, mais selon le degré de précision de l'analyse de ces approches, les éléments identifiés sont plus ou moins détaillés. En outre, selon ce degré de précision également, les approches utilisées pour extraire les éléments des textes sont de complexités variables et peuvent aussi bien être à base d'analyses linguistiques (aspects lexicaux, syntaxiques et sémantiques), structurelles, logiques, ou par modèles de langage. Enfin, il n'existe pas de représentation formelle privilégiée dans l'absolu pour les règles déontiques. L'essentiel est de choisir une représentation qui soit pertinente pour l'application que l'on souhaite en faire. Pour notre cas d'application, une représentation à base de logique du premier ordre serait pertinente.

Remerciements

Ce travail est en partie financé grâce à une bourse de thèse du MESRI attribuée à l'INP-Toulouse, et en partie grâce au projet européen STARLIGHT (oct 2021 - oct 2015), action innovante de l'appel H2020-SU-AI02 - 'Secure and resilient Artificial Intelligence technologies, tools and solutions in support of Law Enforcement and citizen protection, cybersecurity operations and prevention and protection against adversarial Artificial Intelligence'.

Références

- [1] D. Liga and M. Palmirani, Transfer Learning for Deontic Rule Classification : the Case Study of the GDPR, In *International Conference on Legal Knowledge and Information Systems*, Saarbrücken 14-16 December 2022, EasyChair, 2022.
- [2] M. Dragoni, S. Villata, W. Rizzi, and G. Governatori, *Combining NLP Approaches for Rule Extraction from Legal Documents*, Presented at the 1st Workshop on Mining and Reasoning with Legal texts MIREL 2016, 2016. Accessed : Jan. 13, 2023. [Online]. Available :<https://hal.science/hal-01572443>
- [3] A. Wyner and W. Peters, On Rule Extraction from Regulations, In *Frontiers in Artificial Intelligence and Applications*, Vol. 235, Jan. 2011, doi : 10.3233/978-1-60750-981-3-113.
- [4] A. Guissé, F. Lévy, and A. Nazarenko, *Un moteur sémantique pour explorer des textes réglementaires*, IC2011, 2011, Chambéry, France. pp.8. Available :<https://hal.science/hal-00707755>
- [5] T. Mondary, S. Després, A. Nazarenko, and S. Szulman, *Construction d'ontologies à partir de textes : la phase de conceptualisation*, Presented at the 19èmes Journées Francophones d'Ingénierie des Connaissances (IC 2008), Jun. 2008, p. 87. Available : <https://hal.science/hal-00289613>
- [6] G. Ferraro et al., *Automatic Extraction of Legal Norms : Evaluation of Natural Language Processing Tools*, 2020, pp. 64–81. doi : 10.1007/978-3-030-58790-1_5.
- [7] G. Recski, B. Lellmann, A. Kovacs, and A. Hanbury, *Explainable Rule Extraction via Semantic Graphs*, 2021.
- [8] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, A. Paschke, LegalRuleML : XML-Based Rules and Norms, In F. Olken, M. Palmirani, D. Sotara(eds), *Rule-Based Modeling and Computing on the Semantic Web*, RuleML, 2011, Lecture Notes in Computer Science, vol 7018, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-24908-2_30
- [9] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, GATE : an Architecture for Development of Robust HLT applications, In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp 168–175, Philadelphia, 2002, Penn., USA, Association for Computational Linguistics.
- [10] G. Governatori, A. Rotolo, A Conceptually Rich Model of Business Process Compliance, In *Proceedings of the 7th Asia-Pacific Conference on Conceptual Modelling*, pp. 3–12, APCCM 2010, ACS, Brisbane, QLD, Australia, 2010.
- [11] T. F. Gordon, G. Governatori, A. Rotolo, Rules and Norms : Requirements for Rule Interchange Languages in the Legal Domain, In *Rule Interchange and Applications*, G. Governatori, J. Hall, and A. Paschke, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 2009, pp. 282–296, doi : 10.1007/978-3-642-04985-9_26.
- [12] P.-Y. Gicquel, J. Bouché-Pillon, P. Zaraté, N. Aussenac-Gilles, Y. Chevalier, Ontologies and Rules for Access Control : a Feature Oriented Survey, In *1st WS on Collaboration in knowledge discovery and decision making : Applications to sustainable agriculture (DECISIONING 2022)*, Jun 2022, La Plata, Argentina, pp.1-12, (hal-03762626)
- [13] G. Sartor, M. Araszkievicz, K. Atkinson et al., Thirty years of Artificial Intelligence and Law : the second decade, In *Artificial Intelligence and Law*, vol. 30, pp.521–557, 2022, <https://doi.org/10.1007/s10506-022-09326-7>
- [14] S. Villata, M. Araszkievicz, K. Ashley et al., Thirty years of artificial intelligence and law : the third decade, In *Artificial Intelligence and Law*, vol. 30, pp.561–591, 2022, <https://doi.org/10.1007/s10506-022-09327-6>
- [15] P. Amselek, Les fonctions normatives ou catégories modales, In *Philosophiques*, vol. 33, no. 2, pp. 391–418, Nov. 2006, doi : 10.7202/013889ar.