

Chapitre 3

Extraction des données **des** demandes et résultats correspondants

3.1 Introduction

Au cœur de l'analyse des décisions de justice se trouve le concept de demande. Il s'agit d'une réclamation ou requête effectuée par une ou plusieurs parties aux juges. Une partie peut demander des dommages-intérêts en réparation d'un préjudice subi, **ou bien** un divorce, **ou bien** des indemnités auxquelles elle pense avoir droit, ou encore une étude d'expert, etc. Les demandes sont fondamentales car l'argumentation au cours d'une affaire a deux buts : faire accepter ses demandes, et faire rejeter celle de la partie adverse. L'extraction des demandes et des résultats correspondants, dans un corpus, permet ainsi de récolter des données informant de la manière dont sont jugés des types de demandes d'intérêt. Les informations qui nous intéressent sont la catégorie de la demande, le quantum (montant) demandé, le sens du résultat (par ex. la demande a-t-elle été acceptée ou rejetée?), et le quantum obtenu (décidé par les juges). Pour pouvoir extraire les demandes et les résultats, il est nécessaire de comprendre comment ils sont exprimés et co-référencés dans les décisions jurisprudentielles. Leur énoncé peut comporter des expressions plus ou moins **de** complexes, dont souvent des références à des jugements antérieurs, des agrégations ou des restrictions (Figure 3.1).

Jennifer M., Catherine M. et Sandra M. ... demandent à la Cour de :
 - les recevoir régulièrement appelantes incidentes du **jugement du 23/05/2014** ;
 - infirmer **le dit jugement** en **toutes ses dispositions** ; ...
 Statuant à nouveau ...
 - les condamner au paiement d'une somme de 3 000,00 € pour procédure abusive et aux entiers dépens ;

(a) Exemples d'énoncés de demandes

La cour, ...
 CONFIRME **le jugement entreprise** en **toutes ses dispositions**.
 Y ajoutant
 CONSTATE que Amélanie Gitane P. épouse M. est défaillante à rapporter la preuve d'une occupation trentenaire lui permettant d'invoquer la prescription acquisitive de la parcelle BH 377 située [...].
 DEBOUTE Amélanie Gitane P. épouse M. de sa demande en dommages et intérêts.
 CONDAMNE Amélanie Gitane P. épouse M. aux dépens d'appel.
 DIT n'y avoir lieu à l'application de l'article 700 du Code de Procédure Civile.

(b) Exemple d'énoncés de résultats

Figure 3.1 – Énoncés simples, ou comprenant des **références** et des **agrégations** (extraits de la décision 14/01082 de la cour d'appel de Saint-Denis (Réunion))

3.1.1 Données cibles à extraire

3.1.1.1 Catégorie de demande

Une catégorie c de demande regroupe les prétentions qui sont de même nature par le fait qu'elles partagent deux aspects : l'objet demandé (par ex. dommages-intérêts, amende civile, déclaration de créance) et le fondement c'est-à-dire les règles ou normes ou principes juridiques qui fondent la demande (par ex. article 700 du code de procédure civile). Des noms particuliers sont utilisés pour identifier les catégories (Tableau 3.1).

3.1.1.2 Quantum demandé

Le quantum demandé quantifie l'objet de la demande. Nous le notons q_d . Par exemple, dans l'exemple de la Figure 3.1a, "3000 €" est le quantum demandé au titre des dommages-intérêts pour procédure abusive. Bien que, cette étude ne porte que sur des montants d'argent, le quantum peut être d'une autre nature comme par exemple une période dans le temps (garde d'enfant, ou emprisonnement, etc.). Toutes les catégories demandes n'ont

Label	Expression nominative	Objet	Fondement
acpa	amende civile pour abus de procédure	amende civile	Articles 32-1 code de procédure civile + 559 code de procédure civile
concdel	dommages-intérêts pour concurrence déloyale	dommages-intérêts	Article 1382 du code civil
danais	dommages-intérêts pour abus de procédure	dommages-intérêts	Articles 32-1 code de procédure civile + 1382 code de procédure civile
dcppc	déclaration de créance au passif de la procédure collective	déclaration de créance	L622-24 code de commerce
doris	dommages-intérêts pour trouble de voisinage	dommages-intérêts	principe de responsabilité pour trouble anormal de voisinage
styx	frais irrépétibles	dommages-intérêts	Article 700 du code de procédure civile

Les labels ont été définis particulièrement pour cette étude, n'existent pas par conséquent dans le langage juridique.

Tableau 3.1 – Exemples de catégories de demandes

pas **un** quantum (par ex. une demande de divorce) et seul le sens du résultat sera la donnée à extraire dans ce cas.

3.1.1.3 Sens du résultat

Le sens du résultat est l'interprétation de la décision des juges sur une demande. Nous le notons s_r . En général, le sens peut être positif si la demande a été acceptée, et négatif si elle a été rejetée. Il arrive aussi que le résultat soit reporté à un jugement futur ; il s'agit dans ce cas d'un sursis à statuer.

3.1.1.4 Quantum obtenu ou résultat

Le quantum obtenu quantifie le résultat ou la décision des juges. Nous le notons q_r . Il est en général inférieur ou égal au quantum demandé. Si la demande est rejetée, q_r est évidemment nul même si cela n'est pas explicitement mentionné dans le document. Il doit être de la même nature que le quantum demandé (**montant** d'argent ou durée).

3.1.2 Expression, défis et indicateurs d'extraction

Les demandes sont, en général, décrites à la fin de la section d'exposé des faits, procédures, moyens et prétentions des parties (section Litige). Elles rentrent donc dans les "moyens et prétentions des parties" qui regroupent les demandes et les arguments des parties. Quant aux résultats, ils sont décrits dans la section Dispositif et dans la section Motifs (raisonnement des juges). Les demandes sont exprimées en paragraphe où chaque paragraphe correspond soit à une partie, soit à un groupe de partie partageant les mêmes demandes (par ex. des époux). Le paragraphe est parfois organisé en liste dont chaque élément exprime une ou plusieurs demandes, ou fait référence à un jugement antérieur. Les résultats ont aussi la forme de liste dans la section Dispositif. Par contre, dans les motifs de la décision, les raisonnements sont organisés en paragraphes, et ordonnés catégorie après catégorie. Le résultat est donné à la fin du groupe de paragraphes associé à la catégorie.

Cette pseudo-structure n'est pas standard et elle impose de nombreux défis à relever. En effet, une décision jurisprudentielle porte sur plusieurs demandes de catégories différentes ou similaires. Il est important de faire correspondre un quantum demandé extrait au sens et quantum du résultat qui font référence à la même demande. **la** séparation des demandes et des résultats rend difficile cette mise en correspondance. Ce problème peut aussi être causé par la redondance des quanta ; par **ex.** les résultats exprimés dans les Motifs sont résumés dans le Dispositif. D'autre part, les références aux jugements antérieurs exigent de résoudre des références aux résultats de jugements antérieurs qui sont, généralement, rappelés dans le même document. Notons aussi que les difficultés **liés** aux agrégations (par ex. "*infirmer ... en toutes ces dispositions*") et aux restrictions/sélections (par ex. "*infirme le jugement ... sauf en ce qu'il a condamné M. A. ...*") devraient être résolues. Par ailleurs, les catégories de demandes sont nombreuses (500+ code NAC [?]) mais ne sont pas toutes présentes dans toutes les décisions. Tous ces aspects rendent difficiles **l'**annotation manuelle des données de référence et la

modélisation d'une approche d'extraction adéquate. Cependant, nous avons remarqué quelques indicateurs qui pourraient être aidé.

On pourrait au préalable annoter les candidats potentiels de quanta. Nous nous sommes intéressés aux demandes dont les quanta sont des sommes d'argent. Les mentions de somme d'argent sont généralement la forme « [valeur] [monnaie] » (par ex. 3000 €, 15 503 676 francs, un euro, 339.000 XPF). Des centimes apparaissent parfois (par ex. dix huit euros et soixante quatorze centimes, 26'977 € 19). Ainsi, il est possible d'annoter les sommes d'argent à l'aide d'une expression régulière. Même s'il est difficile de reconnaître des sommes d'argent écrites en lettre, il faut remarqué que l'équivalent en chiffre est généralement mentionné tout près (par ex. neuf mille cinq cent soixante six euros et quatre vingt sept centimes (9566,87 €)).

La terminologie utilisée est aussi un bon indicateur pour reconnaître des demandes et des résultats. En effet, le vocabulaire utilisé est très souvent propre aux catégories de demandes. Par exemple le dernier élément de la Figure 3.1a comprend le terme "*pour procédure abusive*" qui est près d'une somme d'argent (3000 €); il est donc probable que ce type de terme assez particulier soit un bon indicateur de la position des quanta. Par ailleurs, des verbes particuliers sont utilisés pour exprimer les demandes et résultats : infirmer, confirmer, constater, débouter, dire, ...

3.1.3 Formulation du problème

Nous avons tenu compte de deux principaux aspects du problèmes :

1. Une décision comprend plusieurs demandes de catégories similaires ou différentes ;
2. Il existe un grand nombre de catégories (500+); ce qui rend difficile l'annotation d'exemples de référence pour couvrir toutes ces catégories.

Nous avons par conséquent opter pour une extraction par catégorie. L'idée est de pouvoir ajouter progressivement de nouvelles catégories. Une exécution du système d'extraction permet ainsi d'extraire les demandes d'une seule catégorie. Le problème est décomposé **grossièrement** en deux principales tâches :

Tâche 1 : Détecter les catégories présentes dans le document pour **n'**appliquer l'extraction **que pour** ces catégories ;

Tâche 2 : Pour chaque catégorie c identifiée, extraire les demandes :

1. identification des valeurs d'attributs : quanta demandés (q_d), quanta obtenus (q_r), et sens du résultat (s_r) ;
2. mise en correspondance des attributs pour former les triplets (q_d, s_r, q_r) correspondants aux paires demande-résultat **de** catégorie c .

3.2 Travaux connexes

Chacune des tâches précédentes se rapproche d'une tâche couramment traité en fouille de texte. En effet, la détection de catégories dans les décisions peut être modélisée comme un problème de classification de document. La tâche d'extraction se rapproche plus des problématiques comme l'extraction d'évènements, le remplissage de champs, ou encore l'extraction de relations et la résolution de référencement.

3.2.1 Problèmes analogues : extraction **de** d'éléments structurés

Les demandes ressemblent aux structures telles que les relations ou les évènements. En effet, les champs définis par ACE [2008], pour les relations, et ACE [2005] pour les évènements, se rapprochent de ceux **que** visés lors de l'extraction des demandes comme l'illustre le Tableau 3.2. Plus précisément, une catégorie de demande **e** correspond à un type d'évènement ou de relation entre deux entités. Les arguments qui participent à l'évènement « demande »

ou à la relation « demande-résultat » sont le quantum demandé et le quantum résultat. Le sens du résultat représente la classe de la structure « demande ».

	Relation [ACE, 2008]	Événement [ACE, 2005]	Analogie chez les demandes
Type	Org-Aff.Student-Alum	Die	Catégorie="Dommages-intérêts pour procédure abusive"
Passage (<i>extend</i>)	<i>Card graduated from the University of South Carolina</i>	"Il est mort hier d'une insuffisance rénale."	(Figure 3.1)
Déclencheur (<i>trigger</i>)	-	"mort"	"procédure abusive"
Participants ou Arguments (<i>arguments</i>)	Arg1="Card" Arg2="the University of South Carolina"	Victim-Arg="il" Time-Arg="hier"	Quantum-demandé="3000€" Quantum-obtenu="0 €"
Classes (<i>attributes, classes</i>)	Asserted	Polarity=POSITIVE, Tense=PAST	Sens-résultat="Rejeté"

Tableau 3.2 – Exemples d’analogie entre relations, événements et demandes

3.2.2 Approches d’extraction d’éléments structurés

L’extraction d’éléments structurés a généralement une formulation modulaire du problème en tâches plus simples. D’une part, on a l’identification des déclencheurs¹ et des arguments. D’autre part, une mise en correspondance relie les arguments et déclencheurs qui participent à la même relation ou au même événement. Les classes peuvent être déterminées par classification du passage associé. Cette décomposition a permis à de nombreuses méthodes de voir le jour.

L’approche traditionnelle consiste à traiter à l’aide d’un processus enchainant des modules adaptés chacun à une tâche. La sortie d’une étape est l’entrée de la suivante. C’est ainsi que Ahn [2006] définit un enchaînement de modèles de classification (k-plus-proches-voisins [Cover and Hart,

1. terme-clé indiquant la présence d’un événement [ACE, 2005].

1967] vs. classificateur d'entropie maximum [Nigam et al., 1999]), pour extraire des champs des événements dans le corpus d'ACE [ACE, 2005]. même si les différents modules sont plus facile à résoudre, ce type d'architecture souffre de l'accumulation et la propagation d'erreurs d'une étape à la suivante, ainsi que de la non exploitation de l'interdépendance entre les tâches. Par conséquent, l'inférence jointe des champs est préconisée. Celle-ci peut-être réalisé par une modélisation graphique probabiliste ou neuronale. Par exemple, pour l'extraction d'évènement, Yang and Mitchell [2016] estiment la probabilité conditionnelle jointe du type d'entité t_i , les rôles des arguments r_i et les types d'entités qui remplissent ces rôles a : $p_\theta(t_i, r_i, a | i, N_i, x)$, i étant un déclencheur candidat, N_i l'ensemble des entités candidates qui sont des potentiels arguments pour i , et x est le document. Par ailleurs, Nguyen et al. [2016] illustrent l'utilisation des réseaux de neurones profonds avec une couche pour la prédiction du déclencheur, une autre pour le rôle des arguments, et la dernière encode la dépendance entre les labels de déclencheurs et les rôles d'arguments.

L'annotation d'ACE [2005] est un marquage des champs dans le texte, et par conséquent, la position ou l'occurrence des champs est indiquée (« annotation au niveau du segment de mot »). Comme dans notre cas, les données peuvent être annotées dans un tableau, hors des textes d'où elles sont issues. Il est donc nécessaire de retrouver leur position sans supervision. Palm et al. [2017] proposent dans cette logique une architecture de réseaux de neurones point-à-point qu'ils ont expérimentés sur des corpus de requêtes de recherche de restaurant et films [Liu et al., 2013] ou de réservation de billets d'avion [Price, 1990]. Ils se sont intéressés au problème de remplissage de champs en apprenant la correspondance entre les textes et les valeurs de sorties. Leur modèle est basé sur les réseaux de pointeurs [Vinyals et al., 2015] qui sont des modèles séquence-à-séquence avec attention, dans lesquelles la sortie est une position de la séquence d'entrée. Le modèle proposé consiste en un encodeur de la phrase et des contextes, plusieurs décodeurs (un pour chaque

champ). L'application de cette architecture à l'extraction des demandes serait confrontée à deux obstacles majeures auxquelles il faut répondre au préalable. Premièrement, les décisions judiciaires ont des contenus de plusieurs centaines à plusieurs milliers de lignes contrairement aux requêtes manipulées par Palm et al. [2017] dont la plus longue ne comprend que quelques dizaines de mots. La complexité des architectures neuronale de TALN augmente rapidement en espace et par conséquent en temps, avec la longueur des documents manipulés en entier. Deuxièmement, nous disposons de très peu de données annotées ; entre 23 et 198 documents annotés dans notre cas contre plusieurs milliers pour les expérimentations de Palm et al. [2017].

L'avantage de l'utilisation des réseaux de neurones vient de leur capacité à apprendre automatiquement des caractéristiques pertinentes contrairement aux modèles probabilistes qui exigent très souvent une ingénierie manuelle des caractéristiques. Par contre, il est beaucoup plus facile d'utiliser les modèles probabilistes sur des corpus de faible taille et de longs textes comme c'est le cas pour notre problème.

3.2.3 Extraction de la terminologie d'un domaine

L'identification des attributs peut être facilitée grâce à leur proximité avec des termes-clés caractéristiques des catégories de demandes au même titre que les « déclencheurs » aident à identifier les événements. Ne disposant pas au préalable de la liste des termes pertinents pour l'extraction des demandes, il est possible de les apprendre. Il existe à cet effet plusieurs métriques statistiques de pondération de termes généralement employées en recherche d'information et en classification de texte comme méthodes la sélection de caractéristiques. Ces métriques sont qualifiées de poids globales car calculées à partir des occurrences dans un corpus, à la différence des poids locaux (Tableau 3.4) calculées à partir des occurrences dans un document. Quelques métriques sont formulées ici en utilisant les notations du le Tableau 3.3 définies pour une base d'apprentissage.

Notation	Description
t	un terme
d	un document
$ t $	longueur de t (nombre de mots)
c	la catégorie (domaine ciblé)
\bar{c}	la classe complémentaire ou négative
D	ensemble global des documents de taille $ D $
D_c	ensemble des documents de c de taille $ D_c $
$D_{\bar{c}}$	ensemble des documents de \bar{c} de taille $ D_{\bar{c}} $
N_t	nombre de documents contenant t
$N_{\bar{t}}$	nombre de documents ne contenant pas t
$N_{t,c}$	nombre de documents de c contenant $t = a$
$N_{\bar{t},c}$	nombre de documents de c ne contenant pas t
$N_{t,\bar{c}}$	nombre de documents de \bar{c} contenant t
$N_{\bar{t},\bar{c}}$	nombre de documents de \bar{c} ne contenant pas t
$DF_{t c}$	proportion de documents contenant t dans le corpus de c ($DF_{t c} = \frac{N_{t,c}}{ D_c }$)
$DF_{c t}$	proportion de documents appartenant à c dans l'ensemble de ceux qui contiennent t

Tableau 3.3 – Notation utilisée pour formuler les métriques

3.2.3.1 Métriques non-supervisées

Les métriques non-supervisées affectent un score à un terme en rapport avec l'importance de ce dernier dans le corpus global D . Parmi ces métriques, on retrouve **par** la fréquence inverse de document (*inverse document frequency*) *idf* [Sparck Jones, 1972] qui, comme ses variantes *pidf* [Wu and Salton, 1981] et *bidf* [Jones et al., 2000], accorde **plus** d'importance aux termes rares. Elles considèrent en fait qu'un terme rare est plus efficace pour la distinction entre des documents. Par conséquent, elles sont efficaces en recherche d'information mais moins indiquées en classification de texte où le but est plutôt de séparer des catégories [Wu et al., 2017]. Elles se formulent comme

suit :

$$idf(t) = \log_2 \left(\frac{N}{N_t} \right), pidf(t) = \log_2 \left(\frac{N}{N_t} - 1 \right), bidf(t) = \log_2 \left(\frac{N_t + 0.5}{N_t + 0.5} \right)$$

Il est possible de prendre explicitement en compte le fait que les termes peuvent comprendre plusieurs mots (n-grammes) et avoir des tailles différentes (nombre de mots). La C-value [Frantzi et al., 2000], par exemple, distingue la fréquence du terme et de ses sous-termes (termes imbriqués) par la formule :

$$C\text{-value}(t) = \begin{cases} \log_2(|t|) \cdot (N_t - \frac{1}{|T_t|} \cdot \sum_{b \in T_t} N_b), & \text{si } t \text{ est imbriqué} \\ \log_2(|t|) \cdot N_t, & \text{sinon,} \end{cases}$$

T_t étant l'ensemble des termes candidats qui contiennent t .

3.2.3.2 Métriques supervisées

Les métriques supervisées mesurent l'information contenu dans les labels des documents de la base d'apprentissage. Pour un terme t , Elles expriment généralement la différence de proportion qui existe entre les occurrences de t dans D_c et ses occurrences dans $D_{\bar{c}}$. Elles sont ainsi mieux adaptées à la distinction entre catégories. Parmi les nombreuses métriques existantes, nous avons expérimenté les suivantes :

La différence de fréquence Δ_{DF} consiste simplement à calculer la différence entre les proportions de documents contenant t respectivement dans c et \bar{c} :

$$\Delta_{DF}(t, c) = DF_{t|c} - DF_{t|\bar{c}}$$

Le gain d'information ig [Yang and Pedersen, 1997] estime la quantité d'information apportée par la présence ou l'absence d'un terme t sur

l'appartenance d'un document à une classe c :

$$ig(t, c) = \frac{N_{t,c}}{N} * \log_2 \left(\frac{N_{t,c}N}{N_t} \right) + \frac{N_{\bar{t},c}}{N} * \log_2 \left(\frac{N_{\bar{t},c}N}{N_{\bar{t}}|D_c|} \right) \\ + \frac{N_{t,\bar{c}}}{N} * \log_2 \left(\frac{N_{t,\bar{c}}N}{N_t|D_{\bar{c}}|} \right) + \frac{N_{\bar{t},\bar{c}}}{N} * \log_2 \left(\frac{N_{\bar{t},\bar{c}}N}{N_{\bar{t}}|D_{\bar{c}}|} \right)$$

La fréquence de pertinence rf [Lan et al., 2009] a comme intuition de considérer que plus la fréquence d'un terme t est élevé dans D_c relativement à sa fréquence dans $D_{\bar{c}}$, plus il contribue à distinguer les documents de c de ceux de \bar{c} . Elle est calculée par la formule :

$$rf(t, c) = \log \left(2 + \frac{N_{t,c}}{\max(1, N_{t,\bar{c}})} \right)$$

Le coefficient du χ^2 [Schütze et al., 1995] estime le manque d'indépendance entre t et c . Par conséquent, une grande valeur de $\chi^2(t, c)$ indique une relation étroite entre t et c . Elle est calculée par la formule :

$$\chi^2(t, c) = \frac{N((N_{t,c}N_{\bar{t},\bar{c}}) - (N_{t,\bar{c}}N_{\bar{t},c}))^2}{N_t N_{\bar{t}} |D_c| |D_{\bar{c}}|}$$

le coefficient de corrélation ngl de Ng, Goh et Low [Ng et al., 1997] est la racine carré positive du χ^2 [Schütze et al., 1995] :

$$ngl(t, c) = \frac{\sqrt{N}((N_{t,c}N_{\bar{t},\bar{c}}) - (N_{t,\bar{c}}N_{\bar{t},c}))}{\sqrt{N_t N_{\bar{t}} |D_c| |D_{\bar{c}}|}}.$$

L'intuition est de ne regarder que les termes qui proviennent de D_c et qui indiquent l'appartenance à c .

Le coefficient gss de Galavotti, Sebastiani, et Simi [Galavotti et al., 2000] est une fonction simplifiée du ngl [Ng et al., 1997] :

$$gss(t, c) = (N_{t,c}N_{\bar{t},\bar{c}}) - (N_{t,\bar{c}}N_{\bar{t},c}).$$

Le facteur N a été éliminé car il est le même pour tous les termes. Le facteur $\sqrt{N_t N_{\bar{t}}}$ est supprimé car il accentue les termes extrêmement rares qui ne sont pas efficaces pour la classification de textes. Le facteur $\sqrt{|D_c| |D_{\bar{c}}|}$ est éliminé car il accentue les catégories extrêmement rares, ce qui tend à réduire l'efficacité micro-moyennée (efficacité calculée globalement sur le corpus de test sans distinction à priori du label des éléments).

Le test de Marascuilo (mar) qui se calcule par la formule :

$$mar(t, c) = \frac{\left(\begin{aligned} &(N_{t,c} - N_t N_{t,c}/N)^2 \\ &+ (N_{t,\bar{c}} - N_t |D_{\bar{c}}|/N)^2 \\ &+ (N_{\bar{t},c} - |D_c| N_{\bar{t}}/N)^2 \\ &+ (N_{\bar{t},\bar{c}} - N_{\bar{t}} |D_{\bar{c}}|/N)^2 \end{aligned} \right)}{N}$$

Si $mar(t, c) \geq 3.84$ on accepte l'hypothèse selon laquelle la proportion de textes pour laquelle t prédit c est significative [?].

La distance de Kullback-Leibler kld [?] se formule comme suit :

$$kld(t, c) = (N_{t,c}/N_t) * \log\left(\frac{N_{t,c} N}{N_t |D_c|}\right)$$

Le « delta lissé d' idf » , $dsidf$ [Paltoglou and Thelwall, 2010], est une version lissée du delta idf ($didf$) de Martineau et al. [2009] ($didf(t, c) = \log_2 \left(\frac{|D_{\bar{c}}| N_{t,c}}{|D_c| N_{t,\bar{c}}} \right)$). $dsidf$ se formule comme suit :

$$dsidf(t, c) = \log_2 \left(\frac{|D_{\bar{c}}| (N_{t,c} + 0.5)}{|D_c| (N_{t,\bar{c}} + 0.5)} \right)$$

Le delta BM25 d' idf , $dbidf$ [Paltoglou and Thelwall, 2010], est une

autre variante plus sophistiquée du *didf* qui se calcule comme suit :

$$dbidf(t, c) = \log_2 \left(\frac{(|D_{\bar{c}}| - N_{t,\bar{c}} + 0.5)(N_{t,c} + 0.5)}{(|D_c| - N_{t,c} + 0.5)(N_{t,\bar{c}} + 0.5)} \right)$$

3.2.3.3 Discussions

A l'exception de la C-value, ces métriques ne tiennent pas explicitement compte de la taille des termes dans les situations où on souhaiterait manipuler des termes de tailles différentes. [Brown, 2013] propose que soit affecté à un n-gramme t le poids $\left(\frac{N_t}{N}\right)^{0.27} * |t|^{0.09}$, une formule obtenue empiriquement pour l'identification du langage d'un document. Par ailleurs, La méthode C-value [Frantzi et al., 2000] propose un produit similaire avec le logarithme de la longueur à la place des puissances. Il est par conséquent évident que le produit lissé de la longueur du terme (puissance ou logarithme) avec les métriques décrites précédemment, permet de booster les longs termes qui, bien que rares, sont très souvent plus pertinents que certains termes plus courts.

3.3 Méthode

3.3.1 Détection des catégories par classification des documents

Étant données l'ensemble $D_{\bar{c}}$ des documents ne comprenant aucune demande de la catégorie d'intérêt c , nous proposons de modéliser la tâche de détection des catégories en une tâche de classification de documents. Pour chaque catégorie c , un modèle de classification binaire est entraîné pour déterminer si un document d contient une demande de la catégorie c . Nous avons particulièrement expérimenté quatre algorithmes traditionnellement utilisés comme approches de base. Il s'agit du Bayésien Naïf [?], de l'arbre de décision

[?], des k -plus-proches-voisins (KNN) [Cover and Hart, 1967], de la machine à vecteurs de support [?]. Les labels utilisés correspondent aux catégories d'intérêt. Par exemple, un document sera labellisé *danais* s'il contient des demandes de dommages-intérêts pour abus de procédure, et *nodanais* sinon. Chaque document d est représenté sous une forme vectorielle du type TF-IDF (*term frequency - inverse document frequency*) proposé par Salton and Buckley [1988] dont chaque dimension k est identifiée par un terme t_k . Le poids $w(t_k, d)$ affecté à ce dernier est le produit normalisé d'un poids global $g(t_k)$ au corpus du mot et d'un poids local $l(t_k, d)$ de t_k dans le document d : $w(t_k, d) = l(t, d) \times g(t) \times nf(d)$, où nf est un facteur de normalisation tel que la norme cosinus $cos(d) = \sqrt{\sum_k (w(t_k, d))^2}$ qui est généralement utilisé.

Description	Formule
Décompte brute du terme [Salton and Buckley, 1988]	$tf(t, d) = \text{nombre d'occurrences de } t \text{ dans } d$
Présence du terme [Salton and Buckley, 1988]	$tp(t, d) = \begin{cases} 1 & , \text{ si } tf(t, d) > 0 \\ 0 & , \text{ sinon} \end{cases}$
Normalisation logarithmique	$logtf(t, d) = 1 + \log(tf(t, d))$
Fréquence augmentée et normalisée du terme [Salton and Buckley, 1988]	$atf(t, d) = k + (1 - k) \frac{tf(t, d)}{\max_{t \in T} tf(t, d)}$
Normalisation basée sur la fréquence moyenne du terme [Manning et al., 2008] (avg représente la moyenne)	$logave(t, d) = \frac{1 + \log tf(t, d)}{1 + \log \text{avg}_{t \in T} tf(t, d)}$

Tableau 3.4 – Métriques locales

Etant donné le grand nombre de métriques de pondération existantes, la métrique choisie est celle qui fournit la meilleure performance sur les données d'apprentissage.

3.3.2 Extraction basée sur la proximité entre sommes d'argent et les termes-clés

Diverses approches d'extractions d'informations existent (section § 3.2.2). Il paraît important de proposer dans un premier temps une approche basique explorant la solvabilité du problème du fait de ses multiples spécificités

dont l'annotation d'une seule catégorie dans un document qui en contient plusieurs, l'annotation dans un tableau et donc à l'extérieur du document, la très faible quantité des données annotées, la multiplicité des demandes et catégorie dans un même document. Par conséquent, nous proposons ici une chaîne d'extraction à base de termes-clés, applicable pour chaque catégorie de demande. Il s'agit d'une approche qui tente de reproduire une lecture naïve du document en se basant sur des expressions couramment employées pour énoncer les demandes et résultats. La méthode consiste en deux phases dont une phase d'apprentissage des termes-clés de la catégorie, à proximité desquels seront identifiés les attributs durant la phase d'application comme l'illustre la Figure 3.2. On remarque en effet que, naïvement, le seul fait que 1500 euros soit aussi proche des termes-clés *amende civile* et *pour procédure abusive* signifie bien qu'il s'agit du quantum demandé comme amende civile pour procédure abusive.

```
" ...
- débouter M. S. de ...
- le condamner à payer une amende civile de 1.500 euros pour procédure abusive ...
- le condamner à payer la somme ..."
```

(a) Extrait original d'un énoncé de demande avant marquage

```
" ...
- débouter M. S. de ...
- le <demande categorie="acpa">condamner à payer une <terme-clef catego-
rie="acpa">amende civile</terme-clef> de <argent> 1.500 euros </argent> <terme-clef
categorie="acpa"> pour procédure abusive</terme-clef> ...
- le</demande> condamner à payer la somme ..."
```

(b) Énoncé, sommes d'argent, et termes-clés marqués

Figure 3.2 – Illustration de la proximité des quantas et termes-clés

3.3.2.1 Pré-traitement

Le pré-traitement est nécessaire pour :

1. sectionner le document comme décrit au chapitre 2 en sections Entête, Litige, Motifs, Dispositif;

2. annoter les sommes d'argent (en chiffre) à l'aide de l'expression régulière « $[0-9]([0-9]|\['\,'.\]|\backslash s)*\backslash s*([Ee]uro[s]\{0,1\}|\text{franc}[s]\{0,1\}|\text{€}|F|XPF|CFP|EUR|EUROS|[i])(|\$)\gg;$
3. annoter les énoncés de demandes et de résultats respectivement dans les sections Litige et Dispositif. Pour cela, les mots introductifs du tableau 3.5 sont employés car ils indiquent le début d'un énoncé indépendamment de la catégorie.

Demande	Résultat (organisé par polarité ou sens)		
	accepte	sursis à statuer	rejette
<i>accorder, admettre, admission, allouer, condamnation, condamner, fixer, laisser, prononcer, ramener, surseoir</i>	<i>accorde, accordons, admet, admettons, alloue, allouons, condamne, condamnons, déclare, déclarons, fixe, fixons, laisse, laissons, prononce, prononçons</i>	<i>réserve, réservons, sursoit, sursoyons</i>	<i>déboute, déboutons, rejette, rejetons</i>

Tableau 3.5 – Mots introduisant les énoncés de demandes et de résultats

La recherche de passages à l'aide listes de termes est une technique souvent utilisée dans les décisions de justice, à l'exemple de Wyner [2010] qui utilise **de** termes similaires à ceux du Tableau 3.5 pour annoter les énoncés de résultats (toute phrase contenant un terme de jugement) : *affirm, grant, deny, reverse, overturn, remand, ...*

3.3.2.2 Apprentissage des termes-clés d'une catégorie

Les termes-clés sont identifiés à l'aide de méthodes statistiques d'extraction ou sélection de terminologie. La base d'apprentissage comprend les corpus D_c et $D_{\bar{c}}$ dont les documents ont été pré-traités. Le processus d'apprentissage des termes se déroule comme suit :

1. Restreindre le contenu de chaque document de D_c à la concaténation des énoncés de demande et résultats contenant des sommes d'argent de valeur égale à celle des quanta annotés.

2. Restreindre le contenu de chaque document de $D_{\bar{c}}$ à la concaténation des énoncés de demande et résultats contenant des sommes d'argent.
3. à l'aide d'une métrique global g , calculer le score des termes du corpus $D_c \cup D_{\bar{c}}$. Ce score est le produit g' de g avec le logarithme de la longueur du terme, pour booster les termes longs : $g'(t, c) = \log_2(|t|) \times g(t, c)$.
4. normaliser les scores en appliquant à chaque score original ($g'(t, c)$) la formule $g'_{norm}(t, c) = \frac{\max_{t_k}(g'(t_k, c)) - g'(t, c)}{\max_{t_k}(g'(t_k, c)) - \min_{t_k}(g'(t_k, c))}$.
5. trier par ordre décroissant des termes ;
6. sélectionner les premiers termes qui obtiennent les performances optimales sur la base d'apprentissage .

3.3.3 Application de l'extraction à de nouveaux documents

A l'aide des termes-clés appris, l'extraction des données de couples demandes-résultats se déroule comme suit :

1. reconnaître et marquer les occurrences des termes dans le document ;
2. extraire les quanta demandés (q_d) et résultats (q_r) à proximité des termes-clés respectivement dans les énoncés de demande et résultat qui contiennent des sommes d'argent et un terme-clé ;
3. le mot introductif de l'énoncé résultat indique le sens du résultat (s_r) tel que catégorisé dans le Tableau 3.5 ;
4. relier les attributs (q_d, s_r, q_r) correspondant à une même pair demande-résultat :
 - (a) former les paires (énoncé de demande, énoncé de résultat) similaire (nous utilisons la métrique de « la plus longue sous-séquence commune » [Bakkellund, 2009])
 - (b) pour chaque paire d'énoncés formée, relier les quanta demandés et quanta résultats par ordre d'occurrence similaire.

3.4 Résultats expérimentaux

Nous analysons ici la capacité de l’approche proposée à reconnaître efficacement les catégories de demandes présentes dans les documents, et à extraire les valeurs des attributs des différentes paire demande-résultats qui y sont exprimées. Il y est discuté les données et métriques d’évaluation employées, ainsi que des résultats expérimentaux observés avec des exemples annotés pour les six catégories du Tableau 3.1.

3.4.1 Données d’évaluation

L’annotation manuelle d’exemples s’effectue pour une catégorie à la fois afin que la tâche soit plus facile pour les experts. Le protocole d’annotation se déroule en étapes :

1. définir une catégorie c par son objet et sa norme juridique ;
2. former un corpus D_c de documents contenant des demandes de c , et un autre $D_{\bar{c}}$ de documents n’en contenant pas ;
3. extraire toutes les demandes de catégories c mentionnées dans D_c , pour annoter les données des paires demande-résultat dans un tableau comme celui illustré par le Tableau 3.6 ;

	A	B	C	D	F	H	L	N
1	IDENTIFICATION DE LA DECISION			DESCRIPTION DE LA PRETENTION			DESCRIPTION DU RESULTAT	
2	Type	Ressort	RG	OBJET	NORME	QUANTUM	RESULTAT	QUANTUM RESULTAT (obtenu)
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
441	CA	Lyon	14/06911	dommages-intérêts	700 Code de Procédure Civile	3,500.00 €	rejette	0.00 €
442	CA	Lyon	14/06911	dommages-intérêts	700 Code de Procédure Civile	2,000.00 €	accepte	1,500.00 €

Les noms des champs sont sur les 2 premières lignes et les demandes sont données en exemple pour la catégorie *dommages-intérêts sur le fondement de l'article 700 du code de procédure civile* (décision 14/06911 de la cour d’appel de Lyon).

Tableau 3.6 – Extrait du tableau d’annotations manuelles des demandes.

Les demandes de la catégories c , présentes dans le corpus d'évaluation $D_c \cup D_{\bar{c}}$, sont considérées toutes annotées dans le tableau. La répartition des documents d'évaluation est donnée par l'histogramme de la Figure 3.3.

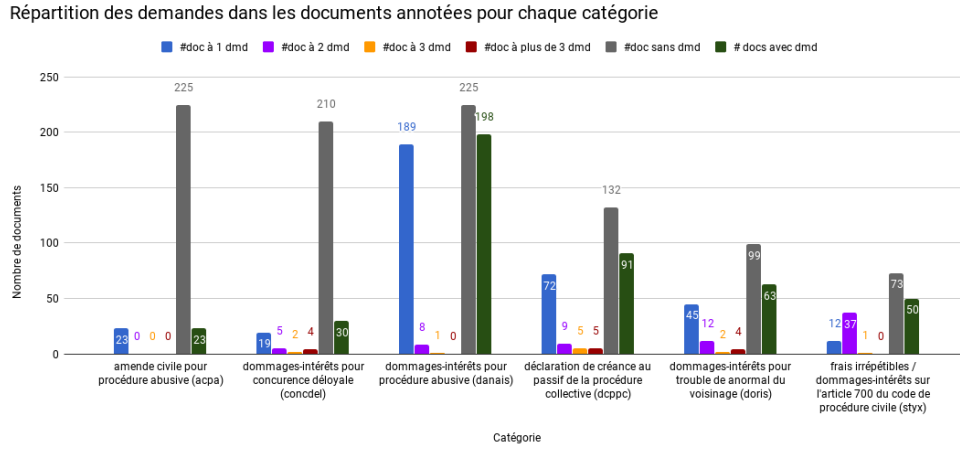


Figure 3.3 – Répartitions des demandes dans les documents annotées.

Il faut aussi noter que bien que l'annotation des demandes et résultat soit réalisée dans un tableau (annotation externe au contenu), elle reste une tâche très difficile. Le très faible nombre de documents annotés manuellement le témoigne. Le nombre maximum de documents annotés pour une catégorie est seulement de 198 (barres vertes de *danais*).

3.4.1.1 Métriques d'évaluation

Reconnaissance de catégorie par classification La classification des documents est évaluée en utilisant les métriques précision (P), rappel (P), f1-mesure (F1) calculées à l'aides des nombres de vrais positifs (TP), faux positifs (FP), faux négatifs (FN) comme suit :

$$P = \frac{TP}{TP + FP}; R = \frac{TP}{TP + FN}; F1 = 2 \times \frac{P \times R}{P + R}$$

Extraction des attributs des paires demande-résultat Nous évaluons les approches proposées sur l'extraction de 3 données : le quantum demandé q_d , le sens du résultat s_r et le quantum obtenu q_r . Une demande est donc un triplet (q_d, s_r, q_r) . Il est possible d'évaluer le système pour un sous-ensemble x de ce triplet sur les demandes extraites d'un corpus annotées D de test. Nous utilisons les métriques traditionnellement employées en extraction d'information : la précision (Eq. 3.1), le rappel (Eq. 3.2), et la F1-mesure (Eq. 3.3).

$$Precision_{c,x,D} = \frac{TP_{c,x,D}}{TP_{c,x,D} + FP_{c,x,D}} \quad (3.1)$$

$$Rappel_{c,x,D} = \frac{TP_{c,x,D}}{TP_{c,x,D} + FN_{c,x,D}} \quad (3.2)$$

$$F1_{c,x,D} = 2 \times \frac{Precision_{c,x,D} \times Rappel_{c,x,D}}{Precision_{c,x,D} + Rappel_{c,x,D}} \quad (3.3)$$

Ces mesures sont définies à partir des nombres de vrais positifs (TP), faux positifs (FP) et faux négatifs (FN). Au niveau d'un document d :

- le nombre de vrais positifs $TP_{c,x,d}$ est le nombre de demandes extraites de d par le système, qui sont effectivement de la catégorie c ;
- le nombre de faux positifs $FP_{c,x,d}$ est le nombre de demandes extraites de d par le système, mais qui ne sont pas des demandes de c (demandes en trop) ;
- le nombre de faux négatifs $FN_{c,x,d}$ est le nombre de demandes annotées comme étant de c mais qui n'ont pas pu être extraites par le système (demandes manquées).

Au niveau d'un corpus d'évaluation D , ces métriques sont sommées :

$$TP_{c,x,D} = \sum_{d \in D} TP_{c,x,d}; FP_{c,x,D} = \sum_{d \in D} FP_{c,x,d}; FN_{c,x,D} = \sum_{d \in D} FN_{c,x,d}$$

Une donnée observée (par exemple « 3 000 € ») est bien extraite auto-

matiquement si sa valeur (le nombre 3000) correspond à celle du quantum annoté dans le tableau. Nous considérons que les unités monétaires, entre les quanta extraits et ceux manuellement annotés, sont égales

3.4.2 Détection des catégories par classification

Les implémentations dans Weka, la bibliothèque JAVA de fouille de données, des modèles de classification ont été utilisées : le classifieur Bayésien naïf (NB), l'arbre de décision (J48), les k-plus-proches-voisins (KNN), et le SVM. A chaque entraînement, s'exécute une sélection de modèle par validation croisée sur les données d'entraînement ; elle a pour but de sélectionner la métrique locale et la métrique globale appropriée. Les résultats obtenus par 5-*folds* validation croisée sont présentés sur le tableau 3.7.

	NB			J48			KNN			SVM		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
acpa	1.0	1.0	1.0	0.996	0.955	0.972	1.0	1.0	1.0	0.996	0.955	0.972
concdel	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.995	0.967	0.979
danais	0.988	0.989	0.988	0.996	0.995	0.995	0.995	0.995	0.995	0.993	0.993	0.993
dcppc	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
doris	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
styx	1.0	1.0	1.0	0.984	0.983	0.983	1.0	1.0	1.0	1.0	1.0	1.0

(P= Précision, R=Rappel, F1 = F1-mesure)

Tableau 3.7 – Résultats d'une 5-*fold* validation croisée pour la détection catégorie

D'après les résultats, la tâche 1 est relativement aisée pour les algorithmes traditionnels qui détectent parfaitement la présence ou non d'une catégorie dans les documents. Par conséquent, pour toute catégorie c , les résultats de l'extraction, dans la suite, ne sont discutés que pour les documents de c , car, grâce à l'efficacité de la phase de classification, aucun document de \bar{c} ne sera traité par la phase d'extraction.

3.4.3 Extraction de données des paires demandes-résultats

Les scores des termes-clés candidats étant normalisés, si on sélectionne les termes dont les scores sont supérieurs à un seuil fixé, on remarque que chaque métriques d'extraction ont un degré d'efficacité dépendant de la catégorie de demande (Tableau 3.8 avec 0.5 comme seuil fixé).

	<i>acpa</i>	<i>concdel</i>	<i>danais</i>	<i>dcppc</i>	<i>doris</i>	<i>styx</i>	Moyenne
<i>bidf</i>	37.33	32.73	23.96	20.46	8.08	28.43	25.17
χ^2	54.55	25.88	43.97	28.35	13.11	52.73	36.43
<i>dbidf</i>	37.58	24.63	56.25	29.06	11.58	52.73	35.31
Δ_{DF}	54.55	25.55	48.16	28.1	19.64	52.73	38.12
<i>dsidf</i>	37.58	25.25	56.42	26.05	8.72	53.46	34.58
<i>gss</i>	54.55	25.11	48.16	28.1	19.64	52.73	38.05
<i>idf</i>	38.78	32.73	22.31	20.53	8.27	25.22	24.64
<i>ig</i>	4	12.4	45.21	14.99	16.74	51.13	24.08
<i>marascuilo</i>	54.55	23.65	43.97	26.67	17.91	52.73	36.58
<i>ngl</i>	42.02	23.97	52.31	27.21	13.29	53.2	35.33
<i>pdf</i>	26.19	33.71	21.83	20.46	8.76	27.68	23.11
<i>rf</i>	41.11	33.09	55.72	28.56	14.93	51.23	37.44

Tableau 3.8 – $F1_{c,(q_d,s_r,q_r),D_c}$ moyenne pour une 5-fold validation croisée pour chaque métrique de sélection de termes pour un seuil égal à 0.5

Par conséquent, la métrique et le seuil adéquats doivent être bien sélectionnés. En choisissant, pour ces méta-paramètres, les valeurs qui donnent les meilleurs performances d'extraction sur la base d'apprentissage, les résultats suivants sont observés (Tableau 3.9).

Ces résultats détaillés font remarquer que les attributs, pris individuellement, présentent d'assez bonnes performances. Cependant, la mise en correspondance des attributs (triplet (q_d, s_r, q_r)) peine toujours à montrer des performances du même rang. On remarque néanmoins que les mesures-F1 (q_d, s_r, q_r) sont proches de celles de l'attribut qui présente le plus de difficulté. L'échec de l'extraction de ces attributs est donc une des causes des performances observées pour la liaison des attributs de paires similaires demande-résultat. Par ailleurs, les données sur le résultat, s_r et q_r , sont en générale plus faciles à extraire que le quantum demandé q_d . Il est aussi bien de noter que la quantité d'exemples annotés de documents ne semble pas être la

<i>c</i>	Données	$ V_c $	Données d'entraînement				Données de test			
			P	R	F1	%Docs	P	R	F1	%Docs
<i>acpa</i>	<i>q_d</i>	1	86.4	56.37	68.13	56.37	68.33	54	58.99	46
	<i>q_r</i>	1	100	65.09	78.74	65.09	93.33	63	71.43	55
	<i>s_r</i>	1	100	65.09	78.74	65.09	93.33	63	71.43	55
	(<i>s_r, q_r</i>)	1	100	65.09	78.74	65.09	93.33	63	71.43	55
	(<i>q_d, s_r, q_r</i>)	1	86.4	56.37	68.13	56.37	68.33	54	58.99	46
<i>concdel</i>	<i>q_d</i>	26	49.33	44.02	45.31	24.17	73.2	29.72	33.29	26.67
	<i>q_r</i>	26	48.3	42.66	44.1	22.5	75.73	28.89	34.3	26.67
	<i>s_r</i>	26	46.52	40.89	42.36	22.5	74.93	26.39	33.09	26.67
	(<i>s_r, q_r</i>)	26	46.52	40.89	42.36	22.5	74.93	26.39	33.09	26.67
	(<i>q_d, s_r, q_r</i>)	26	42.43	37.41	38.68	20.83	68.27	23.06	28.65	23.33
<i>danaïs</i>	<i>q_d</i>	37	77.71	48.71	59.68	37.3	79.25	47.5	59	37.3
	<i>q_r</i>	37	77.68	48.71	59.67	37.03	77.78	46.46	57.79	36.22
	<i>s_r</i>	37	77.05	48.33	59.19	37.03	77.78	46.46	57.79	36.22
	(<i>s_r, q_r</i>)	37	77.05	48.33	59.19	37.03	77.78	46.46	57.79	36.22
	(<i>q_d, s_r, q_r</i>)	37	74.45	46.65	57.16	35.81	74.41	44.38	55.23	34.59
<i>dcppc</i>	<i>q_d</i>	35	45.71	36.64	40.66	34.05	44.64	40.73	41.75	31.4
	<i>q_r</i>	35	78.99	63.21	70.2	59.33	75.48	64.51	68.41	53.82
	<i>s_r</i>	35	84.73	67.85	75.33	63.24	81.21	69.14	73.51	57.43
	(<i>s_r, q_r</i>)	35	78.99	63.21	70.2	59.33	75.48	64.51	68.41	53.82
	(<i>q_d, s_r, q_r</i>)	35	34.2	27.39	30.41	28.03	31.66	28.55	29.41	25.37
<i>doris</i>	<i>q_d</i>	8	31.98	35.76	32.94	7.75	37.48	35.9	36.63	7.12
	<i>q_r</i>	8	35.73	39.72	36.69	8.63	39.43	38.47	38.89	7.12
	<i>s_r</i>	8	35.06	39.56	36.24	9.06	42.91	41.44	42.12	8.94
	(<i>s_r, q_r</i>)	8	32.61	36.16	33.45	8.2	38.14	37.04	37.54	7.12
	(<i>q_d, s_r, q_r</i>)	8	24.48	27.16	25.13	5.61	29.7	28.53	29.08	7.12
<i>styx</i>	<i>q_d</i>	4	69.34	59.55	64.04	33.5	69.3	59.49	63.61	32
	<i>q_r</i>	4	75.87	65.17	70.08	31.5	74.86	64.08	68.63	28
	<i>s_r</i>	4	75.87	65.17	70.08	31.5	74.86	64.08	68.63	28
	(<i>s_r, q_r</i>)	4	75.87	65.17	70.08	31.5	74.86	64.08	68.63	28
	(<i>q_d, s_r, q_r</i>)	4	57.61	49.44	53.19	25.5	57.24	48.36	52.08	24

P : précision, R : rappel, F1 : F1-mesure

%Docs : proportion de documents dont l'ensemble des données extraites est égale à l'attendu (documents parfaitement traités)

 $|V_c|$: nombre moyen de termes-clés identifiés pour la catégorie *c*

Tableau 3.9 – Résultats détaillés pour l'extraction des données avec sélection automatique de la méthode d'extraction des termes-clés

garantie d'une bonne extraction. On remarque en effet que les meilleures performances sont obtenues pour la catégorie disposant du plus faible nombre d'exemples annotés (*acpa*) avec en moyenne un seul terme-clé appris.

3.4.4 Analyse des erreurs

En extraction d'éléments structurés, on retrouve trois types d'erreurs [Yang and Mitchell, 2016] : les données manquées (faux négatifs), les don-

nées en plus (faux positifs), et les mauvaises classifications (confusions). La confusion n'est pas discutée ici car les annotations ne sont faites que pour une seule classe.

Etant donné que la précision est en général supérieure au rappel, il est certain que les erreurs sont majoritairement dues aux données manquées comme le confirme le Tableau 3.10.

	Données d'entraînement		Données de test	
	%erreurs FP	%erreurs FN	%erreurs FP	%erreurs FN
q_d	36.90	63.10	36.52	63.48
q_r	32.30	67.70	34.32	65.68
s_r	31.72	68.28	34.11	65.89
(s_r, q_r)	32.32	67.68	34.39	65.61
(q_d, s_r, q_r)	37.77	62.23	37.72	62.28

Tableau 3.10 – Types et taux d'erreurs (pourcentage en moyenne sur les 6 catégories de demandes)

Trois raisons peuvent expliquer le fait que peu de données attendues soient extraites. Premièrement, certaines valeurs d'attributs ne sont pas mentionnées dans les sections Litige et Dispositif utilisées (pourcentage inférieurs à 100 dans les Tableaux 3.11 et 3.12 comme par exemple les quanta résultat de *doris* plus présents dans la section Motifs que dans le Dispositif).

	# q_d	# $q_d \neq NUL$	# dans doc.	# dans Litige	# dans Motifs	# dans Dispositif
acpa	23	16	16 (100%)	16 (100%)	9 (56.25%)	5 (31.25%)
concdel	58	56	55 (98.21%)	55 (98.21%)	7 (12.5%)	2 (3.57%)
danais	208	182	182 (100%)	179 (100%)	39 (21.43%)	23 (12.64%)
dcppc	126	126	122 (96.83%)	109 (86.51%)	71 (56.35%)	65 (51.59%)
doris	94	83	83 (100%)	82 (98.80%)	21 (25.30%)	6 (7.23%)
styx	89	86	86 (100%)	86 (100%)	12 (13.95%)	9 (10.47%)

Les pourcentages ne sont calculés que pour les valeurs non nulles

Tableau 3.11 – Taux de quanta demandés (q_d) mentionnés dans les documents annotés

Deuxièmement, la sélection des termes-clés n'est pas parfaite (Tableau 3.13). D'une part, l'ensemble sélectionné ne couvre pas toutes les situations d'expression de la catégorie (par exemple, pour la catégorie *styx*, le terme « frais irrépétibles » est souvent utilisés à la place de « article 700 du

	# q_r	# $q_r \neq NUL$	# dans doc.	# dans Litige	# dans Motifs	# dans Dispositif
acpa	23	6	6 (100%)	3 (50%)	6 (100%)	5 (83.33%)
concdel	58	8	8 (100%)	2 (25%)	8 (100%)	6 (75%)
danais	208	23	23 (100%)	15 (65.22%)	22 (95.65%)	20 (86.96%)
dcppc	126	76	75 (98.68%)	55 (72.37%)	56 (73.68%)	64 (84.21%)
doris	94	44	44 (100%)	28 (63.64%)	40 (90.91%)	24 (54.55%)
styx	89	30	29 (96.67%)	16 (53.33%)	22 (73.33%)	29 (96.67%)

Les pourcentages ne sont calculés que pour les valeurs non nulles

Tableau 3.12 – Taux de quanta accordés (q_r) mentionnés dans les documents annotés

code de procédure civile », mais dans très peu d'exemples annotés). D'autre part, certains termes sont trop spécifiques à la base d'apprentissage (par exemple, pour la catégorie *concdel*, des sommes d'argent et autres termes comme « condamner in solidum les sociétés » apparaissent dans la liste).

Catégorie	Termes-clés appris
<i>acpa</i>	amende civile
<i>concdel</i>	titre de la concurrence déloyale, somme de 15000euros à titre, réparation de son préjudice financier, payer la somme de 15000euros, condamner in solidum les sociétés, agissements constitutifs de concurrence déloyale
<i>danais</i>	dommages et intérêts pour procédure, 32-1 du code de procédure, intérêts pour procédure abusive, titre de dommages-intérêts pour procédure, intérêts pour procédure, article 32-1 du code, dommages-intérêts pour procédure abusive
<i>dcppc</i>	admet la créance déclarée, admet la créance, passif de la procédure collective, passif de la procédure, hauteur de la somme, créance déclarée, titre chirographaire, admission de la créance, rejette la créance,
<i>doris</i>	préjudices, abusive, condamner solidairement, solidairement, réparation du préjudice, réparation, titre de dommages et intérêts, dommages, titre de dommages, dommages et intérêts, titre de dommages-intérêts, payer aux époux, jouissance
<i>styx</i>	700 du code de procédure, article 700 du code, 700 du code, article 700, 700

Tableau 3.13 – Premiers termes sélectionnés lors du premier essai de la validation croisée

Troisièmement, les expérimentations ont été réalisées sur des décisions d'appel mais les énoncés, de demande et résultat renvoyant aux décisions de jugements antérieurs, ne sont pas encore traités dans l'approche. Ces références aux décisions antérieures représentent une part importante des demandes discutées dans les décisions d'appel. Il est donc nécessaire de les intégrer explicitement dans le processus d'extraction, pour compléter les données extraites.

3.5 Conclusion

Ce chapitre décrit le problème d'extraction de données pertinentes relatives aux paires demande-résultat mentionnées dans les décisions de justice. Les divers défis relatifs à la tâche y sont discutés en remarquant des analogies avec d'autres tâches classiques de la fouille de données textuelles. Il a été démontré la solvabilité du problème par la proposition et l'expérimentation d'une approche d'extraction basée sur l'apprentissage de la terminologie des catégories de demande et autres connaissances du domaine judiciaire telles que les motifs d'énoncés de demandes et de résultat, ainsi que leur position conventionnelle dans les documents. Les expérimentations démontrent que l'approche permet d'extraire plus ou moins bien des demandes selon la catégorie traitée. A cause de la forte dépendance aux connaissances **à priori**, la méthode rencontre des limites qui ne peuvent être surmontées qu'en rendant la méthode beaucoup plus complexe qu'elle ne l'est déjà. Des approches d'apprentissage automatique sont recommandées comme perspectives. Elles devront être capables d'apprendre l'emplacement des données à extraire de manière semi-supervisée à l'aide de faibles quantités de documents annotés de grande taille.