

# Étude textométrique du profil modal du genre judiciaire (*décision de justice*)

Fadila Taleb

Université Rouen Normandie– fadila.taleb@etu.univ-rouen.fr

## Abstract

In this article, we propose an instrumented analysis of the modal profile (i.e. the study of modes of validation of textual content) of a corpus of court decisions with the implementation of a theoretical model in modal semantics, the *TMM* (Gosselin, 2010). This study, carried out at the micro and meso-textual level with the TXM software, is a preparatory step for a macrotextual analysis of modalities that aims to constitute modal scenarios. The latter are defined as interpretative pathways that account for the textual and argumentative structure of court decisions. They are textual modelling tools intended in this work to help litigants in the reading of court verdict.

**Keywords :** modal profile, modal scenario, court verdict, text zone, modal zone, TMM.

## Résumé

Dans cet article, nous proposons une analyse instrumentée du profil modal (i.e. l'étude des modes de validation du contenu textuel) d'un corpus de décisions de justice avec la mise en œuvre d'un modèle théorique en sémantique des modalités, la *TMM* (Gosselin, 2010). Cette étude, menée au niveau micro et méso-textuelle avec le logiciel TXM, est une étape préparatoire pour une analyse macrotextuelle des modalités qui vise à former des scénarios modaux. Ces derniers sont définis comme des parcours interprétatifs rendant compte de la structure textuelle et argumentative des décisions de justice. Ce sont des outils de modélisation textuelle destinés dans ce travail à aider les justiciables dans la lecture des décisions de justice.

**Mots clés :** profil modal, scénario modal, décision de justice, zone textuelle/zone modale, TMM.

## 1. Introduction

Nous assistons aujourd'hui à une forme d'informatisation accélérée des institutions (privées et publiques), accompagnée d'une production importante de documents numériques. Si les bouleversements promis inquiètent tout autant qu'ils enthousiasment, le besoin d'aider et d'accompagner ce tournant numérique est devenu un enjeu crucial. Dans cette perspective, le projet régional PlaIR<sup>1</sup>, mené à Rouen depuis plusieurs années, regroupe des chercheurs appartenant à des disciplines différentes (informatique, linguistique, droit, histoire, sciences médicales, etc.) et œuvrant ensemble au développement d'outils informatiques qui visent à faciliter l'accès à l'information numérique et numérisée dans trois domaines distincts : la santé, le patrimoine et le droit. Il se situe dans le paradigme de recherche des *humanités numériques* que l'on définit dans un sens très large comme « *un dialogue interdisciplinaire sur la dimension numérique des recherches en sciences humaines et sociales, au niveau des outils, des méthodes, des objets d'études et des modes de communication* » (Dacos & Mounier, 2014, p. 15).

---

<sup>1</sup> Plateforme d'Indexation Régionale : <http://plair.projets.litislab.fr/>

L'étude présentée ici s'inscrit dans ce vaste projet et s'intéresse à la question de l'aide à l'interprétation des documents numériques, plus précisément les décisions de justice. Le besoin pratique d'aider à la lecture de ces textes est formulé par l'Institut du Droit International des Transports (IDIT<sup>2</sup>) dans le but d'améliorer l'accès des adhérents à sa base de données spécialisée, entre autres, dans la recherche d'informations jurisprudentielles. Notre tâche en tant que linguiste consistait en une étude approfondie du fonctionnement discursif du genre judiciaire. L'enjeu de cette description était double. Dans un premier temps, il s'est agi de comprendre les conditions de production et d'interprétation du genre judiciaire et de repérer et cerner ses contraintes interprétatives. Dans un second temps, il a été question de mettre au jour des modèles de scénarios modaux pour trois sous-genres judiciaires : *jugements de tribunaux de commerce (TC)*, *arrêts de cours d'appel (CA)* et *arrêts de la Cour de cassation (CC)*. La notion de scénario modal proposée dans (Gosselin, 2015) comme *un arrière-plan de la signification des lexèmes*<sup>3</sup>, a été redéfinie dans ce travail comme un parcours interprétatif qui rend compte de la structure textuelle et argumentative des décisions de justice. Notre travail repose sur l'hypothèse selon laquelle le scénario modal constitue la clé de voûte de l'interprétation de ces décisions.

Mais avant de reconstituer ces scénarios modaux en procédant à une analyse textuelle des modalités menée directement sur des textes pleins, nous avons commencé par rendre compte du profil modal du corpus (i.e. l'étude des modes de validation du contenu textuel) à travers une étude lexicale menée au niveau micro et mésotextuelle et outillée par le logiciel TXM<sup>4</sup>. C'est cette exploration textométrique du profil modal qui fait l'objet du présent article dont l'objectif principal consiste à démontrer l'intérêt et l'apport d'une analyse menée au niveau du palier micro-linguistique (qui aborde le corpus à travers les entrées lexicogrammaticales) comme étape préalable à une analyse macro-linguistique (structures textuelles plus complexes).

Pour ce faire, notre contribution sera organisée comme suit. Dans la section 2, nous présentons très rapidement le modèle théorique que nous avons choisi pour l'étude des modalités. S'en suivra la section 3 dédiée à la description, typologique et quantitative du corpus d'étude ainsi que les étapes de sa constitution et son encodage, avant de terminer par l'objet central de l'article, l'exploration textométrique du profil modal (section 4).

## 2. La Théorie Modulaire des Modalités (TMM)

S'inscrivant dans la lignée des travaux de Brunot (1922) et Bally (1932), Gosselin (2010) aborde la modalité dans une conception large<sup>5</sup>. Partant, l'un des principes fondateurs de la TMM, est de considérer que chaque énoncé est affecté par une modalité. Celle-ci est définie comme « *un mode de validation/invalidation du contenu représenté* » (Gosselin, 2010, p. 49). Par validation, l'auteur désigne l'opération à travers laquelle une représentation prédiquée est présentée comme valide ou invalide par un énoncé. Tel est l'exemple de la représentation

---

<sup>2</sup> <https://www.idit.fr/>

<sup>3</sup> À titre d'exemple, le scénario modal qui décrit l'arrière-plan du sémantisme du nom *vengeance* se caractérise par l'enchaînement d'événements affectés de modalités appréciatives négatives (Gosselin, 2015, pp. 28-29).

<sup>4</sup> Voir (Heiden, S. et al., 2010).

<sup>5</sup> Les différentes propositions théoriques autour de la modalité sont généralement classées en deux principales conceptions : *étroite (ou restreinte)* et *large*. Dans la première sont regroupées toutes les approches inspirées de la logique modale d'Aristote et réduisant la modalité aux notions de *nécessaire et possible* et à la descriptions de quelques marqueurs modaux. Dans la seconde, sont réunis les travaux définissant la modalité, globalement, comme *l'attitude du locuteur vis-à-vis de son énoncé*. La TMM s'inscrit dans cette dernière conception, mais ne retient pas la définition de la modalité comme *attitude propositionnelle*.

prédiquée « *gendarme (x)* » qu'un locuteur peut présenter comme étant valide : « *Un gendarme a reconnu Paul* », ou invalide : « *ce n'est pas un gendarme qui . . .* », (Ibid. 53). La validation est susceptible de nombreuses variations en fonction de *sa nature objective ou subjective (collective, individuelle), du degré de la prise en charge énonciative* etc. Ainsi, un énoncé, extrait de notre corpus d'étude, comme : « *Selon le rapport de l'expert judiciaire, le bac a joué un rôle causal prépondérant dans la survenance du sinistre* », est présenté sous un mode de validation objectif dans la mesure où son instance de validation, désignée par l'adverbial cadratif « *selon le rapport judiciaire* », renvoie à un fait indépendant de tout jugement subjectif du locuteur. Signalons que l'adverbial cadratif indique également que le locuteur se dissocie de ce mode de validation, mais sans prendre position. Un autre énoncé, tel que : « *La société X prétend ensuite que la cessation des relations commerciales avec la société Y est intervenue après une négociation des tarifs infructueuse* », est affecté d'un mode de validation subjectif, i.e. qu'il met en œuvre une instance de validation de nature subjective. Elle renvoie à « *la société X* ». Remarquons que, contrairement au premier exemple, le locuteur manifeste un refus de prise en charge sur la modalité exprimée dans l'énoncé. Ce refus est indiqué par le verbe *prétendre*.

Afin de cerner dans sa globalité le phénomène hétérogène de la modalité et de décrire tous les aspects (*syntactiques, sémantiques, logiques, énonciatifs*, etc.) qu'implique l'opération de validation, l'auteur de la TMM propose de l'appréhender à travers la combinaison de neuf paramètres qu'il regroupe en trois ensembles: 1) les paramètres conceptuels (instance de validation (I), direction d'ajustement (D) et la force de la validation (F)) qui permettent de définir la catégorie modale (*aléthique, épistémique, appréciative*, etc.) ainsi que sa valeur (*nécessité, certain, interdit*, etc.) ; 2) les paramètres fonctionnels (Niveau dans la hiérarchie syntaxique (N), Portée dans la structure logique (P), Engagement du locuteur (E), Relativité (R) et Temporalité (T)) qui rendent compte du fonctionnement des modalités dans le discours ; 3) le métaparamètre (M) qui décrit la façon dont la valeur des modalités et celle des autres paramètres a été calculée (par marquage linguistique ou inférence contextuelle). Sans aller dans le détail de chacun de ces paramètres, retenons que la combinaison des premiers paramètres conceptuels (instance de validation (I) et direction d'ajustement (D), aboutit à la définition des six catégories modales du français, à savoir : l'aléthique, l'épistémique, l'appréciative, l'axiologique, le boulique et le déontique.

Les modalités aléthiques correspondent aux jugements présentés dans le discours comme objectifs. Ils sont descriptifs et ne véhiculent aucun jugement subjectif, comme dans l'exemple suivant :

- (1) Le conteneur a été chargé à bord du navire Santa Catarina selon connaissance émis le 30 août 2012 à Buenos Aires avec comme chargeur la société Air Liquide Argentina [...].

Les modalités épistémiques sont celles qui se rapportent à la vérité subjective. Elles recouvrent le domaine de la croyance, du savoir et du doute :

- (2) [...] **il ne faisait aucun doute** que Maddens Consulting, agissant pour le compte de Crop's, avait alors transféré à Cna les droits à agir de Crop's en réparation des avaries causées aux marchandises litigieuses alors qu'elles étaient transportées par Hapag Lloyd.

Les modalités appréciatives relèvent, tout comme les modalités épistémiques, de la subjectivité. Cependant, à l'inverse des modalités épistémiques, la catégorie des appréciatives exprime des jugements de valeur (affectifs ou esthétiques) sur le monde. Ces jugements peuvent être collectifs ou individuels :

- (3) Elle a mené une forte action commerciale, qui n'a **malheureusement** donné aucun résultat.

Les modalités axiologiques sont porteuses de jugements de valeurs relatifs à un système

de règles morale, idéologique ou institutionnelle :

- (4) Ces opérations, réalisées selon AHAC de manière **inappropriée et téméraire**, engagent la responsabilité de FRET SA [...].

Les modalités bouliques expriment le souhait, la volonté ou encore le désir. Elles portent essentiellement sur les événements à venir et sont donc particulièrement prospectives :

- (5) la société Sramag n'a pas choisi cette 3ème option, marquant ainsi **sa volonté** d'une arrivée à une date lui laissant un plus long délai de préparation.

Les modalités déontiques renvoient à l'obligatoire, le permis, l'interdit et le facultatif. En s'appuyant sur des conventions institutionnelles, ces valeurs modales sont essentiellement injonctives et s'avèrent donc essentielles pour tout système de normes car elles permettent l'accomplissement des actes illocutoires directifs :

- (6) **il est interdit** à un transporteur de livrer la marchandise sans que lui soit présenté l'original du connaissance.

### 3. Le corpus

#### 3.1. Description typologique et quantitative

Si l'on admet avec (Rastier & Pincemin, 1999, p. 101) que « *l'interprétation est une activité, et non une représentation ou une contemplation, [qui] procède principalement par contextualisation* », il convient aussi d'admettre que le corpus constitue un mode de contextualisation à échelle multiple (Valette, 2016) pour les textes étudiés. En tant que tel, le corpus ne correspond pas au seul groupement de textes construit selon un (ou des) critère(s) spécifiques, mais à plusieurs groupements qui s'emboîtent, les uns relatifs aux autres et ayant chacun un rôle déterminé (Rastier & Pincemin, 1999). De ce point de vue, la typologie des corpus en présence proposée par (Bommier-Pincemin, 1999) et reprise par (Rastier, 2011) nous paraît indispensable pour rendre compte de la complexité de la constitution du corpus qui, rappelons-le, n'est pas un donné ou un préexistant, mais bien un objet construit pour les besoins d'une analyse déterminée. Partant, nous avons proposé de distinguer quatre niveaux différents de corpus : *l'archive, le corpus de référence, le corpus d'étude et les sous-corpus*. Le corpus de référence correspond à la jurisprudence relevant du domaine du droit des transports. Cette dernière est accessible à partir de la base de données du droit international des transports (l'IDIT) qui constitue, quant à elle, l'archive ou le corpus existant. Au sein du corpus de référence, nous avons fixé un nombre limité de décisions, environ 300, retenues pour l'analyse linguistique. Il s'agit du corpus d'étude. Ce dernier a été ensuite divisé en trois sous-corpus pour des analyses comparatives entre les trois sous-genres de la décision : *jugements de tribunaux de commerce (TC), arrêts de la Cour d'appel (CA) et arrêts de la Cour de cassation (CC)*. Chacun des textes composant le corpus d'étude a été également divisé en sous-parties. Ces dernières permettront une description et une caractérisation plus fine et plus approfondie des sous-corpus. Elles sont en nombre de 4 : *les faits, les motifs, les moyens, les dispositifs*.

Le volume textuel du corpus d'étude est de 1014141 occurrences dont 32519 formes. Les deux tableaux suivants 1 et 2 présentent les caractéristiques quantitatives (nombre de textes, nombre de mots (occurrences et formes)) de chaque sous-genre et de chaque partition :

Sous-corpus	Jugements TC	Arrêts CA	Arrêts CC
Caractéristiques			
Nombre de textes	100	100	100
Nombre d'occurrences	364320	407915	241906
Nombre de formes	10304	13222	8993

Tableau 1 Description quantitative des trois sous-corpus

Partitions	Faits	Motifs	Moyens	Dispositifs
Caractéristiques				
Nombre d'occurrences	181509	393092	349433	41846
Nombre de formes	9330	14742	10001	3217

Tableau 2 Description quantitative des sous-parties

### 3.2. Constitution, encodage et prétraitement

Afin de pouvoir exploiter pleinement notre corpus, nous avons eu recours au logiciel TXM. En amont du traitement textométrique, une préparation du corpus était nécessaire. Notre corpus présente l'avantage d'être accessible en ligne, sur la base de données de l'IDIT. Cependant, il convient de préciser que le fonds jurisprudentiel de l'IDIT regroupe des décisions récentes sous format numérique (PDF) facilement exploitable, et des décisions anciennes, numérisées comme images sous format PDF. Ces dernières ont nécessité un traitement d'océrisation (*Reconnaissance Optique de Caractères*) lors de la conversion en textes bruts. Pour constituer le corpus, nous avons opéré en trois principales étapes. Dans un premier temps, et étant donné que le logiciel TXM ne prend pas en charge le format PDF, nous avons procédé à la conversion des textes en format texte brut (.txt). Dans un deuxième temps, et dans le but d'explorer notre corpus à différentes échelles et dans la perspective de mener des analyses contrastives et différentielles, nous avons effectué un balisage avec le langage XML. La structuration du corpus avec des balises XML permet de documenter le corpus et d'enrichir les textes avec des informations multiples, allant d'un niveau micro-linguistique avec un étiquetage grammatical et syntaxique, jusqu'à des renseignements d'un niveau plus global, sur des indications spatio-temporelles, identification des locuteurs, des personnages etc. L'intérêt de recourir à ce langage réside aussi et surtout dans la possibilité de personnaliser les balises et de les adapter en fonction des besoins de l'analyste. Le premier niveau correspond à des balises simples, ne contenant aucun attribut. Ce sont des balises générales, communes à tous les textes du corpus. Elles encadrent les informations relatives au sous-genre (jugement TC, arrêt CA, arrêt CC), à la date du procès, au lieu du procès (uniquement dans les jugements TC et les arrêts CA), à l'identification des parties en procès (appelante et intimée), du magistrat et du greffier. Le deuxième niveau de balisage est plus complexe que le premier. Il intervient sur la structure interne des textes qu'il divise en plusieurs sous-parties. Le découpage s'est effectué en fonction des différentes parties constituant le corps de la décision : *informations, faits, motifs, moyens, dispositif*. Selon le

sous-genre, les divisions divergent<sup>6</sup>. Enfin, dans un troisième temps, nous avons importé le corpus sur le logiciel TXM, sous le format XML/w + CSV. Lors de cet import, TXM a effectué une opération de lemmatisation<sup>7</sup> et un étiquetage morpho-syntaxique (identification de la nature grammaticale) sur les données textuelles grâce à l’outil intégré *Treetagger*<sup>8</sup>.

## 4. Description du profil modal

L’exploration lexicale du corpus a été abordée sous deux angles différents. Dans un premier temps (§ 4.1), pour un aperçu global du corpus, nous sommes partie d’un simple inventaire du lexique organisé hiérarchiquement selon la fréquence (F) d’emploi des mots dans le corpus. Nous nous sommes focalisée ici sur les noms en retenant l’unité linguistique « lemme ». Dans un second temps (§ 4.2) et en plus de l’observation fréquentielle, nous avons utilisé le calcul des spécificités<sup>9</sup> pour mettre en évidence les caractéristiques lexicales des sous-corpus (selon les sous-genres et les zones textuelles).

### 4.1. Approche du corpus par étude fréquentielle (description du corpus selon les sous-genres)

Jugements_TC	F	Arrêts_CA	F	Arrêts_CC	F
société	2273	société	3899	société	2601
article	948	article	919	article	1084
somme	732	transport	723	transport	1013
demande	604	somme	703	marchandise	837
tribunal	480	marchandise	642	appel	782
titre	447	euro	632	code	672
Monsieur	436	demande	596	arrêt	651
marchandise	434	code	581	cour	623
transport	409	procédure	529	contrat	549
responsabilité	395	titre	527	transporteur	539
navire	377	responsabilité	461	faute	397
dépens	360	appel	431	responsabilité	341
compagnie	350	jugement	417	droit	339
France	340	disposition	380	dommage	328
intérêt	340	droit	373	moyen	306
contrat	329	préjudice	372	partie	300
Rouen	327	contrat	355	livraison	292
droit	321	France	350	lieu	285
dommage	314	avocat	337	commerce	280
conteneur	300	application	320	Cour	276
AXA	287	intérêt	306	action	273
cause	284	fait	272	convention	272
partie	253	faute	263	demande	271
Havre	243	dommage	260	somme	269
transporteur	237	barreau	256	décision	251
assureur	236	garantie	253	assureur	247
conclusion	235	cour	246	titre	218
action	232	cause	237	procédure	215
assurance	226	condition	237	préjudice	214
SA	217	paiement	234	prescription	209

Tableau 3 Les 30 premiers substantifs (en lemmes) dans les trois sous-corpus, classés par ordre fréquentiel

<sup>6</sup> À titre d’exemple, si les parties *informations*, *motifs* et *dispositif* sont communes à toutes les décisions, *les faits* sont uniquement présents dans les jugements TC et les arrêts CA et absents dans les arrêts CC.

<sup>7</sup> « La lemmatisation est une opération linguistique qui consiste à ramener les unités graphiques (notamment toutes les flexions) à leur unité de sens c’est-à-dire aux lemmes (les formes canoniques qui servent d’entrées dans les dictionnaires). Ainsi, dans la phrase « je suis parti », le mot graphique « parti » sera ramené à son lemme *Partir* (verbe) lorsque le même mot graphique dans la phrase « vive le parti ! » sera identifié au lemme *Parti* (nom) » (Mayaffre, 2009, p. 4).

<sup>8</sup> *TreeTagger* est un outil d’annotation qui a été conçu par Helmut Schmid dans le cadre du « TC project », à l’Institut de linguistique computationnelle de l’université de Stuttgart. Il peut être utilisé pour annoter des textes de différentes langues, parmi lesquelles : l’allemands, l’anglais, le français, le danois, le portugais, le chinois, le swahili, etc.

<sup>9</sup> Le calcul des spécificités implémenté dans TXM repose sur la loi hypergéométrique développée par (Lafon, 1980). Sans entrer dans le formalisme mathématique de ce calcul 1, disons simplement qu’il offre la possibilité de voir, statistiquement, les mots sur-employés et sous-employés dans les différentes zones textuelles ou dans les sous-corpus (par rapport au corpus global).

Le tableau 3 ci-dessus, énumère les 30 lexèmes les plus fréquents dans chacun des trois sous-corpus (*TC, CA, CC*). Une première observation globale du tableau attire immédiatement l'attention sur la répétition quasi identique, avec un ordre parfois différent, des mêmes termes dans les trois sous-corpus. Il nous semble que cela est notamment dû à l'homogénéité générique des sous-corpus qui relèvent tous d'un même genre textuel lié à une même pratique sociale qui impose clairement ses contraintes lexicales et détermine le choix des vocables utilisés. Ce constat ne nous semble pas inattendu et est généralement commun aux études statistiques menées sur des corpus homogènes et clos<sup>10</sup> et rejoint l'idée selon laquelle le genre informe et détermine la composante lexicale des textes. Par ailleurs, s'agissant de trois sous-genres et donc d'une « lignée de réécriture » (Rastier & Pincemin, 1999), l'ensemble des textes étudiés ne sont que des « générations » dans cette lignée, i.e. des reprises, des transformations, des suites des uns et des autres. Dans les termes de (Genette, 1982, p. 11), cela renvoie au phénomène de l'hypertextualité qu'il a relevé dans les genres littéraires et qu'il décrit comme « *toute relation unissant un texte B (hypertexte) à un texte antérieur A (hypotexte) sur lequel il se greffe d'une manière qui n'est pas celle du commentaire* ». En effet, dans notre cas, un arrêt de la Cour d'appel n'a aucune existence sans un premier jugement antérieur, ici celui du tribunal de commerce, et un arrêt de la Cour de cassation ne peut être sans qu'il ait eu un jugement TC ou un arrêt CA.

Nonobstant le fort rapprochement lexical entre les trois sous-corpus, il semble, tout de même, que l'usage de certains mots reste spécifique à l'un ou à l'autre. Ainsi, nous remarquons que dans les jugements TC, il y a davantage de lexique axé sur la description des faits et relevant de la langue courante (*navire, conteneur, palette, date, France, Havre etc.*) lorsque dans les arrêts CA et CC le vocabulaire dominant relève du langage juridique, celui de la norme (*code, procédure, contrat, disposition, avocat, droit etc.*). Ce constat est nécessairement lié aux différents statuts juridiques attribués aux trois institutions judiciaires. Le tribunal de commerce étant la première juridiction à prendre connaissance du litige accorde une place importante dans les jugements, plus que les deux autres juridictions, à la description des événements à l'origine de la contestation. En Cour d'appel, les faits sont certes réexaminés, étudiés et rapportés dans la décision, mais son rôle consiste également à vérifier l'appréciation retenue par les premiers juges et leur application du droit. Dans ce sens, et dans une mesure beaucoup moins importante, la mission de la Cour d'appel se rapproche de celle de la Cour de la cassation et cela justifie la proximité lexicale encore plus saillante entre ces deux sous-genres, comparés aux jugements TC.

Du point de vue du profil modal, l'observation du tableau révèle, dans les trois sous-corpus, la dominance de trois principales catégories modales au sein desquelles nous pouvons ranger les items lexicaux. La première, manifestant une forte présence est la modalité aléthique qui regroupe des lexèmes comme : *transporteur Tribunal, navire, marchandise, conteneur, contrat, cour, France, Rouen, AXA, Havre, barreau, avocat, etc.* Leur présence ici témoigne de la particularité du genre jurisprudentiel qui, contrairement au discours législatif par exemple, rend compte de l'application concrète du droit à des cas d'espèce, des situations réelles que le juge décrit dans les jugements et les arrêts. La deuxième est la modalité déontique, très propice au genre judiciaire qui relève d'un discours fondamentalement déontique. Cette catégorie regroupe des lexèmes comme *droit, obligation*, qui sont des modalités dénotées<sup>11</sup>, et d'autres auxquels la modalité est associée<sup>12</sup> au niveau sublexical dans

<sup>10</sup> Dans *Paroles de Présidents* (Mayaffre, 2004) a relevé cette influence générique sur la composante lexicale des textes. En confrontant le discours du président Jacques Chirac aux discours de ses prédécesseurs, les présidents sous la Ve République (de de Gaulle à Mitterrand), il a ainsi observé que les 10 premiers mots les plus utilisés par tous les présidents étaient quasiment identiques et reflètent l'« architecture immuable du discours présidentiel » (p. 24).

<sup>11</sup> Les modalités dénotées sont portées par un nombre limité de lexèmes qui « expriment directement des modes de validation susceptibles de porter sur divers contenus » (Gosselin, 2010, p. 103).

<sup>12</sup> Les modalités associées sont portées par tous les lexèmes. Elles peuvent être linguistiquement marquée ou

le sens où elle est marquée par l'un des termes contenus dans leur définition (*responsabilité*<sup>13</sup>, *procédure*<sup>14</sup>, *loi*<sup>15</sup>, etc. Enfin, la troisième catégorie, également caractéristique du discours juridique dans la mesure où ce dernier se fonde principalement sur des jugements de valeur moraux et légaux, est la modalité axiologique. Les lexèmes relevés réfèrent soit à des normes idéologiques générales, ou à des normes et valeur spécifiques au droit des transports. Dans cette catégorie, on trouve des lexèmes comme *faute*, *vol*, *respect*, *justice*, *équité*, *dol*, *responsabilité*, *vice*, *infraction* qui sont porteurs de modalités axiologiques absolues dans le sens où ils expriment directement le louable ou le blâmable. D'autres lexèmes expriment des modalités axiologiques négatives relativement aux conséquences blâmables ou louables qu'ils induisent. Ces conséquences apparaissant généralement au niveau sublexical, les modalités qui leur sont associées sont dites *associées au niveau sublexical*. Dans cette catégorie, citons l'exemple de *dommage*<sup>16</sup>, *préjudice*<sup>17</sup>, ou encore *avarie*<sup>18</sup>.

Cette exploration globale nous permet de dessiner un premier profil modal du genre judiciaire qui se caractérise par trois principales catégories modales : *aléthique*, *axiologique* et *déontique*. Pour compléter cette analyse et aller vers une caractérisation plus fine, selon des variables précises, nous allons dans la section suivante, nous intéresser aux spécificités des zones textuelles.

#### 4.2. Approche du corpus par le calcul des spécificités (description des zones textuelles)

Faits		Motifs		Moyens		Conclusions	
Formes	S+	Formes	S+	Formes	S+	Formes	S+
2002	30	Attendu	30	ALORS	30	Ainsi	30
société	29,4	Considérant	30	AUX	30	ANNULE	30
mars	23,1	Mais	30	Elle	30	audience	30
2009	18,3	n'	30	MOYEN	30	cassation	30
2000	17,2	pas	30	MOYENS	30	CASSE	30
assigné	17	Qu'	30	n'	30	cassé	30
Havre	14,9	responsabilité	30	QUE	30	civile	30
destination	14,4	Sur	30	faute	28,2	condamne	30
février	14,3	selon	28,6	attaqué	17,5	Condamne	30
assigner	13,1	sera	23,4	subsidaire	16,2	Confirme	30
confié	12,7	ne	19,8	attaquée	16,2	Déboute	30
Suivant	12,6	condamnera	18,3	elle	16	demandes	30
acte	12,3	succombe	17,5	hypothèse	15,6	dépens	30
adresse	12,1	SUR	16,3	Moyens	15,4	Dit	30
Monsieur	11,9	inéquitable	15,8	PREMIER	15	financière	30
dénommée	11,6	donc	15,3	ANNEXES	13,8	jugé	30
2001	10	comme	15,3	violé	13,6	jugement	30
juillet	9,7	convient	13,6	FAIT	11,3	liquidés	30
mai	9,5	obligations	12,6	GRIEF	11,3	Ordonne	30
2004	9,2	alors	11,9	juger	11	payer	30
septembre	8,9	condamnée	10,7	SECOND	10,9	prononcé	30
août	8,9	équité	10,3	écritures	10,7	Reçoit	30
2010	8,9	code	8,9	dire	10,6	rejette	30
port	8,5	déboutera	8,5	lourde	10,6	Rejette	29,1
palette	8,4	appréciations	7,9	constater	9,8	euros	28,4
cartons	8,2	justifie	7,4	légale	9,8	Déclare	26,5
expertise	8	déclarera	6,8	juges	7,8	publiquement	25,9
RALSTON	7,9	existence	6,7	privé	7,6	Statuant	25,7
GUYOT	7,9	inopérantes	6	valoir	6,8	renvoie	25,7
octobre	7,6	justement	5,6	débouter	6,7	fondées	21,8

Tableau 4 Les 30 mots (formes graphiques) sur-utilisées selon les trois zones textuelles (faits, motifs, dispositifs)

pragmatiquement inférées.

<sup>13</sup> « Obligation de réparer un dommage causé » (Cabrillac, 2017, p. 485).

<sup>14</sup> « Manière de procéder juridiquement, ensemble de règles suivant lesquelles doivent se dérouler les actions en justice » (TLFi).

<sup>15</sup> « Règle, prescription émanant de l'autorité souveraine dans une société donnée et entraînant pour tous les individus l'obligation de s'y soumettre sous peine de sanctions. » (TLFi).

<sup>16</sup> « Atteinte subie par une personne » (Cabrillac, 2017, p. 208), « Ce qui résulte d'une action qui nuit » (TLFi)

<sup>17</sup> « Acte ou événement le plus souvent contraire au droit ou à la justice, nuisible aux intérêts de quelqu'un » (TLFi).

<sup>18</sup> « Dommage, perte ou dépense extraordinaire survenant au cours d'une expédition maritime et touchant le navire ou la cargaison » (TLFi).



Dans le tableau ci-dessus, nous reprenons les 30 mots (en forme graphique) caractéristiques de chaque zone textuelle par rapport à l'ensemble du corpus. D'abord, leur observation locale permettra de définir la thématique, le profil modal (en se basant sur le lexique) et le rôle de chacune d'elles dans le corps de la décision. Ensuite, étendue à un niveau global, elle servira à déceler les éventuelles particularités selon les sous-genres.

#### 4.2.1. La zone des faits

Dans la zone des faits, nous observons une forte présence d'entités nommées dont des indications spatio-temporelles : *2002, 2009, 2000, mars, février, juillet, date, Havre, Chine, Pologne, Djakarta*, des noms propres (de société, d'experts, d'avocats) : *RALSTON, GUYOT, HAVEN, AXA, DESMARAIS, VISCOUNT*, ou encore des noms d'objets : *conteneur, palette, cartons*, ou des noms désignant des événements ou des actions : *collision, assignation, expertise, abordage*, etc. Inversement, les mots sous-employés dans cette partie de texte, sont ceux que l'on trouve sur-utilisés dans les autres zones textuelles. À titre d'exemple, le graphe exposé dans la figure 1 donne à voir les 10 spécificités négatives (S-) de la zone des faits comparativement à un sous-corpus constitué des autres zones textuelles (*désormais ZT*). Parmi les mots dont l'indice de spécificité est négatif, nous avons relevé dans la zone des faits les adverbes de négation : *ne* (-30), *pas* (-18), le connecteur *Attendu* (-17.3), les lexèmes *faute* (-9), *obligation* (-7.3), *responsabilité* (-6.5), etc.

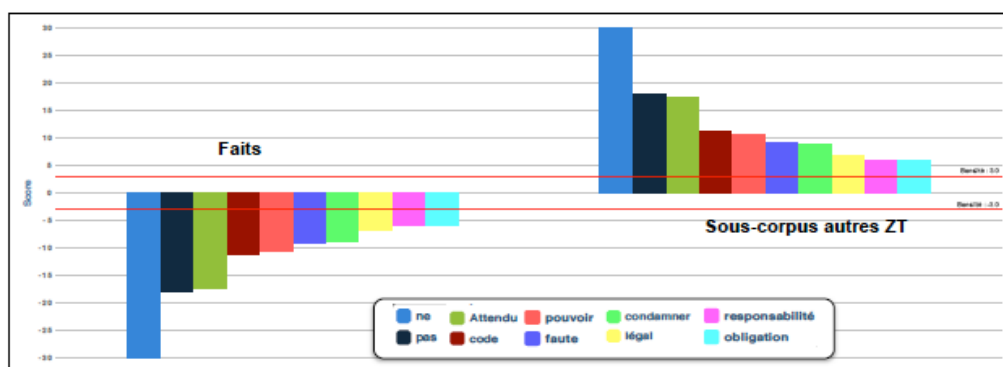


Figure 1 Exemples des spécificités négatives relevées dans la zone des faits comparativement aux autres ZT

La comparaison des (S+) exposées dans le tableau et les (S-) représentées dans cette figure dressent d'une part un profil modal particulier pour la zone des faits et permettent, d'autre part, de caractériser son contenu sémantique par rapport aux autres zones. Du point de vue modal, toutes les unités ont en commun une caractéristique saillante relative à leur mode de validation qui ne véhicule ni un jugement subjectif ni un jugement de valeur. Le recensement de tout le lexique de la zone des faits a fait apparaître la prégnance des lexèmes porteurs de modalités aléthiques. Il convient de préciser deux choses importantes au sujet de ces modalités. D'abord qu'il s'agit là, dans cette première esquisse du profil modal du corpus, de valeurs intrinsèques associées, linguistiquement marquées. Selon les contextes, d'autres modalités peuvent être inférées. Concernant cette prégnance des modalités aléthiques dans la zone des faits, elle témoigne du rôle que celle-ci occupe dans le corps de la décision, i.e. une zone dédiée au rappel des circonstances du litige. Les jugements exprimés ici sont donc principalement descriptifs et tendent vers un degré maximal d'objectivité.

### 4.3. La zone des motifs

Dans la zone des motifs, en revanche, nous constatons dans le tableau 4, une présence notable de modalités axiologiques portées par des lexèmes tels que *condamner, inéquitable, équité, juste, légal, faute, responsabilité, dommage, préjudice*, etc., et des modalités déontiques exprimées par *devoir, condamner, obligation, droit, interdire (interdiction), permettre, falloir*<sup>19</sup>, etc. Contrairement à la zone des faits, le calcul des spécificités sur le corpus des motifs a révélé un déficit important de cette zone en lexèmes porteurs de modalités aléthiques. Le graphe exposé ci-dessous (figure 2) illustre les 10 premiers mots sous-employés dans la zone des motifs comparativement au sous-corpus des autres ZT.

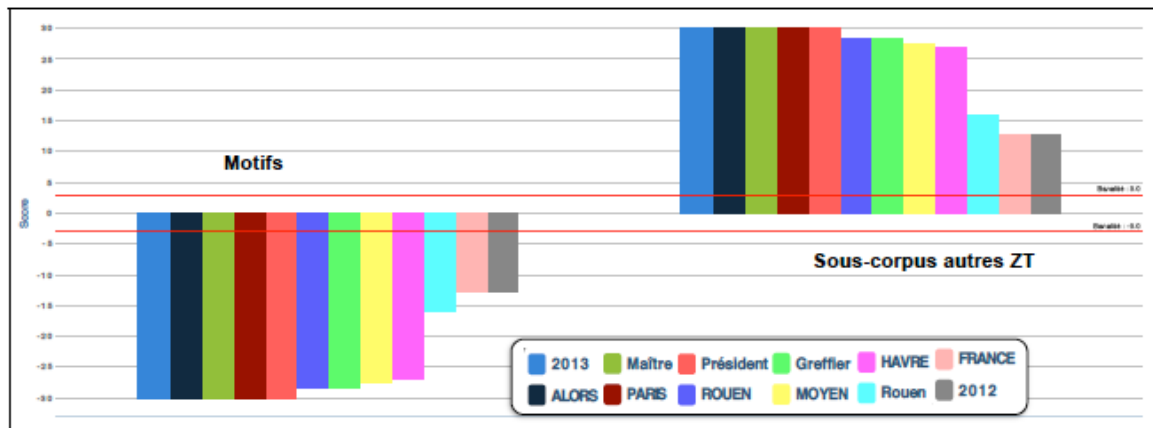


Figure 2 Exemples des spécificités négatives relevées dans la zone des motifs comparativement aux autres ZT

À l'instar des exemples donnés dans cette figure, nous avons relevé que le lexique le moins utilisé dans cette partie du texte est porteur de modalités aléthique, comme les indications spatio-temporelles *2013 (-30)*, *Havre (-26.76)* etc. Ces observations lexicales permettent de postuler que le profil modal de la zone des motifs s'oppose principalement à celui de la zone des faits. En outre, comparativement à la zone précédente, nous remarquons la présence significative d'adverbes, de conjonctions et de prépositions. Cette zone paraît plus riche et syntaxiquement plus complexe. Ces deux constats rendent compte de la particularité de cette partie du texte principalement dédiée à l'exposé de l'argumentation du juge. Ce dernier est en effet dans l'obligation de justifier sa décision en vertu de l'article 455 du code de procédure civile qui postule, on ne peut plus clairement, que « le jugement doit être motivé ». La motivation est donc le cœur de la décision, le lieu du débat, de la confrontation des faits avec le droit. Cela favorise le recours aux jugements de valeur. Cette partie se caractérise donc par une abondance d'expressions et de marqueurs modaux porteurs de jugements subjectifs et évaluatifs : *bien fondé, mal fondé, à juste titre*, etc. Tous ces indices sont importants à relever et à signaler dans le cadre de l'aide à l'interprétation, car ils constituent des moments clés dans l'argumentation judiciaire dans la mesure où ils renseignent sur le type de décision visée par le juge.

### 4.4. La zone du dispositif

Dans la zone du dispositif, il ressort un usage particulier de verbes performatifs. Dans les termes de J.L. Austin (1970), il s'agit plus exactement d'une zone remplie d'actes *verdictifs*<sup>20</sup>

<sup>19</sup> Dans une construction impersonnelle comme dans l'exemple « ... il faut, outre le paiement effectué par l'assureur, que l'indemnité soit due ce qui implique que la responsabilité civile de l'assuré existe ».

<sup>20</sup> « [...] énonciations qui consistent à exprimer ce que l'on a constaté (officiellement ou non), à partir de l'évidence ou des

à travers lesquels le juge formule la réponse effective, la solution finale que les parties attendent. Le magistrat *Condamne, Dit, Déboute, rejette, ANNULE, CASSE, CONFIRME*, etc. Du point de vue modal, ces performatifs expriment des modalités déontiques à travers lesquelles le juge somme les parties d'obligations à respecter. Ces dernières sont généralement des obligations de paiement, ce qui explique l'usage notoire et abondant des sémèmes appartenant à ce champs sémantique : *somme, dépens, payer, financières, euros*, etc. Les modalités déontiques marquent le résultat du processus argumentatif et la fin du litige. Ce

sont également elles qui dotent la décision d'un principe de sacralité, ce que l'on appelle dans le langage juridique *l'autorité (ou le pouvoir ou la force) de la chose jugée*<sup>21</sup>. Cette zone textuelle est la zone du déontique. Comme pour les zones précédentes, nous avons également relevé les spécificités négatives. Soit la figure 3 suivante :

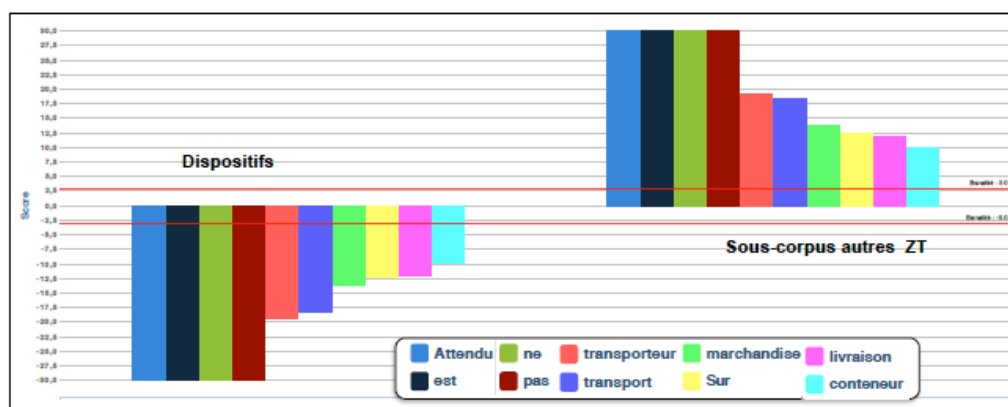


Figure 3 Exemples des spécificités négatives relevées dans la zone des dispositifs comparativement aux autres ZT

Le lexique sous-employé dans cette partie du texte est d'une part le lexique caractérisant la partie argumentative de la décision (adverbes de négation *ne* (-30), *pas* (-30) ; connecteur *Attendu* (-30)) ; et d'autre part, celui qui est sur-utilisé dans la partie narrative du texte (*transporteur* (-19.27), *marchandise* (-13.74), *conteneur* (-9.84), etc.). Ces statistiques confirment le rôle de la zone du dispositif dans le corps de la décision. Elle vient clôturer le mouvement argumentatif exposé dans la zone des motifs et annoncer la solution retenue par le juge au litige opposant les parties en procès.

## 5. Conclusion

Les analyses textométriques proposées dans cet article visaient une caractérisation du profil modal du corpus en guise d'introduction à une étude macrotextuelle sur les scénarios modaux. Dans notre perspective textuelle, cette investigation statistique a été abordée dans une démarche contrastive, d'abord entre les trois sous-genres (TC, CA, CC), puis entre les différentes zones textuelles (faits, motifs, moyens, dispositifs). Nous avons, en outre, pris le soin de proposer deux approches statistiques complémentaires : une approche fréquentielle et une approche par le calcul de spécificités. Alors que l'une nous permettait une description

*raisons concernant les faits ou leur caractère axiologique. Il s'agit d'actes judiciaires, plutôt que législatifs ou exécutifs* » (Austin, 1970 : 182).

<sup>21</sup> Selon l'article 480 du code de procédure civile : « *Le jugement qui tranche dans son dispositif tout ou partie du principal, ou celui qui statue sur une exception de procédure, une fin de non-recevoir ou tout autre incident a, dès son prononcé, l'autorité de la chose jugée relativement à la contestation qu'il tranche* ».

globale du corpus, l'autre nous offrait une caractérisation plus fine des différents sous-corpus. Cette étude menée au niveau micro-méso-textuelle a montré une richesse du corpus, dans son ensemble, en trois catégories modales : *aléthique*, *axiologique* et *déontique*. Si les deux dernières semblent inhérentes au genre judiciaire qui relève d'un discours institutionnel fondé sur des systèmes de valeur, la première s'est justifiée par la particularité des textes étudiés qui rendaient compte de l'application du droit à des cas d'espèce. Par ailleurs, l'étude contrastive a révélé que ces trois catégories modales se recoupaient avec le découpage des textes en zones textuelles. Ainsi, l'inventaire des spécificités lexicales de chacune des zones et leur analyse modale nous ont fait ressortir trois principales zones modales : *zone aléthique*, *zone axiologique* et *zone déontique*. Elles correspondent, respectivement, à la zone des faits, des motifs et des dispositifs. Cela nous a conduit à retenir la définition de la zone modale comme *un passage textuel caractérisé par une catégorie modale dominante*. L'introduction de cette notion est indispensable pour l'étude macro-textuelle des modalités et pour la description des scénarios modaux.

## Bibliographie

- AUSTIN, J. L. (1970). *Quand dire, c'est faire*, trad. Gilles Lane, Paris : Seuil.
- BALLY, C. 1932, (éd. 1965). *Linguistique générale et linguistique française*, (édition revue et corrigée), Bern : Francke.
- BRUNOT, F. 1922, (éd. 1936). *La pensée et la langue*, Paris : Masson.
- BOMMIER-PINCEMIN, B. (1999). *Diffusion ciblée automatique d'informations. Conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, Thèse de doctorat sous la direction de François Rastier, Université Paris IV Sorbonne.
- CABRILLAC, R. (dir.). (2017). *Dictionnaire du vocabulaire juridique*. Paris : LexisNexis.
- DACOS, M., et MOUNIER, P. 2014. « Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international », *Institut français/ministère des Affaires étrangères pour l'action culturelle*, [En ligne] Consulté le 10 août 2019. URL : <https://www.enssib.fr/bibliotheque-numerique/documents/65357-humanitesnumeriques-etat-des-lieux-et-positionnement-de-la-recherche-francaise-dans-lecontexte-international.pdf>
- GENETTE, G. (1982). *Palimpsestes. La littérature au second degré*, Paris : Seuil.
- GOSELIN, L. (2010). *Les modalités en français. La validation des représentations*, Amsterdam-New York : Rodopi.
- 2015. « De l'opposition modus/dictum à la distinction entre modalités extrinsèques et modalités intrinsèques », *Bulletin de la Société de Linguistique de Paris*, CX/1, pp. 1–50.
- HEIDEN, S., LAFON, P., ILLOUZ, G., HABERT, B., FLEURY, S., & FOLCH, H. (1999). « Maîtriser les déluges de données hétérogènes », in *Actes Corpus et traitement automatique des langues : pour une réflexion méthodologique*, Cargèse, 12–17 juillet 1999 pp. 37–46 [En ligne] Consulté le 25 octobre 2016. URL : <https://halshs.archives-ouvertes.fr/halshs-00151841/document>
- HEIDEN, S., MAGUÉ, J.-Ph. & PINCEMIN, B. (2010). « Txm : Une plateforme logicielle open-source pour la textométrie-conception et développement », in *10th International Conference on the Statistical Analysis of Textual Data (JADT)*, Milano : Edizioni Universitarie di Lettere Economia Diritto, pp. 1021–1032 [En ligne] Consulté le 4 juillet 2016. URL : <https://halshs.archivesouvertes.fr/halshs-00549779/document>
- LAFON, P. (1980). « Sur la variabilité de la fréquence des formes dans un corpus », *Mots. Les langages du politique* 1/1, pp. 127–165.
- MAYAFFRE, D. (2004). *Paroles de Présidents. Jacques Chirac (1995-2003) et le discours présidentiel sous la Ve République*, Paris : Honoré Champion.
- (2009). « L'analyse du discours assistée par ordinateur », [En ligne] Consulté le 23 janvier 2016. URL [http://eprints.aidenligne.francasuniversite.auf.org/19/1/pdf\\_Formation\\_Mayaffre\\_Alexandrie\\_dec09\\_.pdf](http://eprints.aidenligne.francasuniversite.auf.org/19/1/pdf_Formation_Mayaffre_Alexandrie_dec09_.pdf)
- RASTIER, F. & PINCEMIN, B. (1999). « Des genres à l'intertexte », *Cahiers de praxématique* 33, pp. 83–111.
- RASTIER, F. (2011). *La mesure et le grain. Sémantique de corpus*, Paris : Champion.
- VALETTE, M. (2009). *Approche textuelle du lexique, Mémoire Habilitation à diriger des recherches, Institut national des langues et civilisations orientales (INALCO)*.
- (2016). « Analyse statistique des données textuelles et traitement automatique des langues. Une étude comparée », in *International Conference on Statistical Analysis of Textual Data (JADT)*, pp. 697–706 [En ligne] Consulté le 15 décembre 2019 URL : <https://hal-inalco.archives-ouvertes.fr/hal-01335084/document>