



HAL
open science

**Le style clair en droit : étude comparative du discours
juridique en anglais et en français, avant et après
simplification. Le plain language dans la communication
juridique avec le grand public.**

Manon Bouye

► **To cite this version:**

Manon Bouye. Le style clair en droit : étude comparative du discours juridique en anglais et en français, avant et après simplification. Le plain language dans la communication juridique avec le grand public.. Linguistique. Université Paris Cité, 2022. Français. NNT : 2022UNIP7055 . tel-04210917

HAL Id: tel-04210917

<https://theses.hal.science/tel-04210917v1>

Submitted on 19 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris Cité

Ecole Doctorale 622 Sciences du Langage

UR 3967 – CLILLAC-ARP

(Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de
Corpus-Atelier de Recherche sur la Parole)

Le style clair en droit : étude comparative du discours juridique en anglais et en français, avant et après simplification.

*Le plain language dans la communication juridique avec le grand
public*

Par Manon Bouyé

Thèse de doctorat en linguistique anglaise

Dirigée par Monsieur le Professeur Christopher Gledhill

Présentée et soutenue publiquement le 12 septembre 2022

Devant un jury composé de

Monsieur Nicolas Ballier, Professeur, Université Paris Cité, Examineur

Madame Anne Condamines, Directrice de recherche, Université Toulouse-Jean Jaurès,
Examinatrice

Monsieur Laurent Gautier, Professeur, Université de Bourgogne, Rapporteur

Monsieur Christopher Gledhill, Professeur, Université Paris Cité, Directeur de thèse

Madame Susan Hunston, Professeure, University of Birmingham, Examinatrice

Monsieur Julien Longhi, Professeur, Cergy Paris Université, Rapporteur

Résumé

Titre : Le style clair en droit : étude comparative du discours juridique en anglais et en français, avant et après simplification. Le *plain language* dans la communication juridique avec le grand public.

La complexité du discours du droit a souvent été soulignée par les spécialistes du droit et les associations de consommateurs dans les pays anglophones et francophones. Ce constat a mené à la publication de recommandations pour encourager les expert.es du droit à mettre en œuvre un style clair (plain language), en suivant une série de règles rédactionnelles, qui incluent par exemple d'utiliser un vocabulaire courant, de faire des phrases courtes ou d'éviter le passif. Le but de ces recommandations est d'améliorer la lisibilité des documents juridiques pour les non-spécialistes. L'objectif de cette thèse est d'étudier la reformulation du discours juridique par les acteurs institutionnels et associatifs afin que les justiciables puissent avoir accès à leurs droits et prendre des décisions éclairées. Pour cela, nous comparons un corpus de textes anglophones et francophones intitulé LEX, composé de textes juridiques réputés hautement complexes, à un corpus bilingue intitulé PLAIN, constitué de textes qui se réclament d'un style clair, et destinés à un lectorat non-expert. Ces corpus contiennent plusieurs genres discursifs issus de différents pays (Canada, Royaume-Uni, Nouvelle-Zélande et France) et institutions.

Nous proposons une méthode systématique pour caractériser les textes qui s'affichent comme appliquant les recommandations rédactionnelles pour un style clair dans la communication juridique. Ce faisant, nous interrogeons également la conceptualisation et les mesures de la complexité textuelle en linguistique. Adoptant une approche descriptive de l'analyse de textes, nous empruntons à divers cadres théoriques, de la Grammaire Systémique Fonctionnelle à l'analyse de discours et au Traitement Automatique des Langues. Nous caractérisons le passage du corpus LEX à PLAIN en combinant des analyses qualitatives avec des méthodes quantitatives. Ces dernières incluent une classification automatique de textes fondée sur des métriques

de complexité, l'analyse de la terminologie et du lexique ainsi que des schémas phraséologiques liés au passif et des procès sémantiques.

Les deux premières parties de cette thèse présentent le cadre théorique, les corpus utilisés et la méthodologie. Dans la troisième partie, nous confrontons les textes de diffusion du discours du droit aux recommandations pour un style clair. Nos résultats montrent que les recommandations ne sont pas toujours suivies, et quand c'est le cas, nous tentons d'expliquer pourquoi. Les métriques de complexité confirment le pouvoir de prédiction de certains scores classiques de lisibilité pour classer les textes selon leur degré de spécialisation, et révèlent également une plus grande complexité verbale dans les textes dits simplifiés. D'après nos résultats, la densité d'éléments lexicaux jugés complexes diminue lors du passage du registre juridique au style clair, mais pas celle des termes juridiques. Les analyses du passif suggèrent que celui-ci remplit des fonctions discursives précises dans le discours du droit, mais aussi dans sa version simplifiée. Dans nos analyses des stratégies rhétoriques de reformulation du discours du droit, nous nous concentrons en particulier sur les différences et similarités liées au genre discursif ou à la langue (anglais ou français). L'étude des schémas phraséologiques et des procès sémantiques révèle une voix discursive différente dans la communication avec les justiciables non-expert selon le contexte culturel, avec une préférence pour un ton procédural et soutenu en français, par opposition à une plus forte dimension interpersonnelle et une forme d'oralité en anglais.

Mots-clés : anglais de spécialité, français, analyse de discours, linguistique de corpus comparée, *plain language*, complexité, droit, médiation juridique, phraséologie

Abstract

Title : Plain language and the law : a comparative study of English and French legal discourse before and after simplification. Plain language in legal communication with the general public.

The complexity of legal language has often been pointed out by consumer organizations and legal experts alike in English-speaking as well as French-speaking countries. This has led to the publication of recommendations for expert legal drafters to implement plain language by following a number of rules that include using "familiar" words and short sentences, or avoiding the passive. The overall goal of such recommendations is to make legal documents more readable for non-experts. The aim of this thesis is to study how the language of legal texts is reformulated by organizations and government agencies, so that non-expert citizens can know their rights and make informed decisions. To do this, we compare a corpus of English-speaking and French-speaking texts considered to be highly technical and complex (the expert-written corpus called LEX), with a corpus of texts that are claimed to express legal concepts and processes clearly for a non-expert readership (called PLAIN). These corpora comprise various genres from institutions in various countries (Canada, United Kingdom, New Zealand and France).

We set out a systematic method to characterize the texts that are claimed to have undergone simplification following plain language guidelines. By doing so, we also question the notion of textual complexity and how it is measured and conceptualized in linguistics. Our study draws on various theoretical frameworks and tools from corpus linguistics, Systemic Functional Grammar and Natural Language Processing and adopts a descriptive approach to text analysis. To characterize the passage from LEX to PLAIN corpora, we use both qualitative and quantitative methodologies. The latter include supervised learning classification techniques based on complexity metrics, as well as the quantitative analysis of terminology, lexico-grammatical patterns associated with the passive and semantic processes.

The first two parts of this thesis present the theoretical framework, our corpora and the methods used. In a third part, we confront our corpora with plain language recommendations and we describe some of the discourse strategies used in the dissemination of legal texts. Our results suggest that the recommendations are not always followed, and we try to account for this. Complexity metrics computed on our corpora confirm the predictive power of some to distinguish legal texts from plain texts, and reveal that plain texts are characterized by more verbal complexity. Based on our results, the density of so-called complex lexical items decreases in plain texts, but specialized terms do not. Results concerning the passive suggest that it has discursive functions both in legal and plain genres. In our analyses, we pay particular attention to divergent or similar textual features at macro- and micro-textual levels across genres and languages (French and English). The study of phraseological patterns and semantic processes reveals a different textual voice when addressing non-expert readers according to the cultural context, with a preference for a formal, procedural tone in French as opposed to a stronger interpersonal dimension and orality in English.

Keywords : discourse analysis, comparative corpus linguistics, English for Specific purposes, French, complexity, plain language, legal language, dissemination, phraseology

Remerciements

Cette thèse fut un long voyage, et je dois beaucoup à de nombreuses personnes. En premier lieu, je souhaite exprimer ma profonde gratitude à Christopher Gledhill pour avoir accepté de m'encadrer depuis le Master 2. Je ne saurais trop le remercier pour sa patience, et pour les nombreux conseils dont il m'a fait profiter et qui ont été riches d'enseignements. Sa rigueur, ses réflexions et relectures ont façonné ce travail.

Je remercie sincèrement Laurent Gautier et Julien Longhi d'avoir aimablement accepté d'être pré-rapporteurs de cette thèse, ainsi que Susan Hunston, Nicolas Ballier et Anne Condamines de m'avoir fait l'honneur de l'examiner. Je remercie ces deux derniers d'avoir accepté de faire partie de mon comité de suivi, et ainsi guidé ma réflexion au cours de ce travail.

Je tiens à remercier également Nicolas Ballier et Thomas Gaillat de m'avoir donné la chance de participer au projet Ulysses 2019 et d'avoir répondu à mes questions lors de nombreux et fructueux échanges. Cette thèse ne serait pas ce qu'elle est sans eux et sans le travail de l'équipe de l'Insight Centre for Data Analytics à Galway : Andrew Simpkin, Bernardo Stearns, Annanda Sousa, et Manel Zarrouk. Merci à eux et elles de m'avoir accueillie et ouvert les portes du *machine-learning*.

Au sein de l'université, je remercie chaleureusement les équipes de l'ED 622 et du CLILLAC-ARP pour leur accueil, en particulier les docteur.es et futur.es docteur.es Hanna Martikainen, Sylvain Navarro, Claire Kloppmann, Célia Atzeni, et Margaux Guillerit, pour leur soutien et les bons moments passés à Olympe de Gouges, ainsi que Marina Pietri, du Pôle de santé au travail. A l'ENS Paris-Saclay, je remercie mes collègues du département des Langues pour leur compréhension et leur soutien dans les derniers mois de cette thèse.

Je souhaite également témoigner de ma profonde reconnaissance envers toutes les personnes qui m'ont accordé leur temps et leur amitié : Catherine Colin, pour ses encouragements et son aide dans l'organisation de la rédaction ; Jonathan Vacher, qui

m'a très aimablement fourni son code et patiemment expliqué (autant de fois qu'il a fallu) ce qu'est une régression logistique ; Quentin Dénigot, pour son aide technique, sa disponibilité sans faille, et moult boissons chaudes partagées ; Bernard Saintourens, qui m'a fourni de précieuses références et a été mon guide à la faculté de droit de Bordeaux ; Claire O'Riordan et Adriano Ferraresi, pour avoir répondu à mes questions et grandement guidé mes choix méthodologiques ; Kenan Kergrene, pour son expertise LateX quasi-professionnelle ; Anthéa Guittard, Laura-May Simard, Yoann Micheli et Clara Barféty, pour leurs relectures attentives et pour m'avoir écoutée me plaindre pendant près de cinq ans ; toutes celles et ceux qui m'ont accompagnée et soutenue pendant mes études : Sarah Bocquel, Lucile Audiot, Elise Amfreville, Swann Marx, et Alice Hammel.

J'adresse tout particulièrement mes remerciements à Alisson Poisson, dont l'amitié m'a suivie tout au long de mon parcours d'angliciste et qui a généreusement partagé avec moi son expertise de greffière, et à la Dr Alice Brunet, qui, par son expérience de jeune chercheuse, ses précieuses relectures et ses encouragements constants, surtout pendant la phase de rédaction, a véritablement été ma boussole et mon phare dans la nuit.

Pour son soutien indéfectible, je remercie du fond du cœur Baptiste "Ron" Durand, qui m'a épaulée et vaillamment supportée. Merci de m'avoir prêté ses talents de programmation pour m'aider à constituer mon corpus et dessiner de belles figures, et surtout d'avoir pris soin de moi dans la fameuse dernière ligne droite. Je remercie ma grand-mère, ma sœur, mon frère et ma nièce pour leur soutien moral ou matériel. Enfin, toute ma gratitude va à ceux qui ont fait naître en moi le goût de la recherche, à qui je dédie ce travail : mon père, qui m'a toujours encouragée et écoutée parler de cette thèse avec intérêt (en conférence ou ailleurs), et ma mère, pour le modèle qu'elle est pour moi et pour ses relectures et conseils avisés. *Thank you for your love and support.*

Liste des abréviations

Abréviation	Signification
CAB	Citizens Advice Bureau
CSC	Cour Suprême du Canada
DILA	Direction de l'Information Légale et Administrative
GN	Groupe nominal
GP	Groupe prépositionnel
GSF	Grammaire Systémique Fonctionnelle
GV	Groupe verbal
MLC	<i>Mean Length of Clause</i> Longueur Moyenne des Propositions
MLS	<i>Mean Length of Sentence</i> Longueur Moyenne des Phrases
MLT	<i>Mean Length of T-unit</i> Longueur Moyenne des T-units
PL	<i>Plain Language</i> : style clair
PLM	Plain Language Movement
pmm	Par million de mots (traduction de <i>per million tokens</i>)
Schéma LG	Schéma lexico-grammatical
TAL	Traitement Automatique des Langues

Table des matières

Introduction	24
0.1 Genèse du projet	26
0.2 Problématique et objectifs	28
I Cadre théorique	31
1 Le discours juridique et sa diffusion	33
1.1 Le domaine du droit et son étude linguistique	33
1.1.1 La langue et le droit	34
1.1.2 Familles juridiques et territorialité du droit	35
1.1.3 Le droit : un domaine spécialisé hétérogène	36
1.1.4 Différentes approches des discours du droit	39
1.2 Typologie des discours du droit étudiés	41
1.2.1 Le registre juridique	41
1.2.2 Le discours législatif	43
1.2.3 Le discours judiciaire et juridictionnel	45
1.2.4 Le discours administratif	48
1.3 <i>How To Speak Legalese</i> : Traits linguistiques et rhétoriques du dis- cours du droit	49

TABLE DES MATIÈRES

1.3.1	Perceptions et réputation du langage du droit	50
1.3.2	La terminologie et le lexique juridiques, une clé vers le discours du droit	52
1.3.3	Caractéristiques discursives du discours du droit	57
1.3.4	La phraséologie du droit : collocations juridiques et formulaïcité	64
1.4	Diffuser le discours du droit : comment ouvrir le domaine spécialisé aux justiciables ?	67
1.4.1	Des acteurs divers	68
1.4.2	Fonctions et genres de la diffusion du droit	70
1.4.3	Hétérogénéité des lectorats visés	73
1.4.4	Stratégies de reformulation et d’explicitation des discours spé- cialisés	75
1.5	Résumé du Chapitre 1	77
2	Le style clair en droit	79
2.1	Le style clair : définitions et champs d’application	79
2.1.1	<i>Plain language</i> et autres langues simplifiées	80
2.1.2	Le style clair, un principe de rédaction qui touche à différents domaines	82
2.2	Le style clair en droit : histoire et développement	84
2.2.1	Développement historique du <i>plain language</i> en droit dans les pays anglo-saxons	85
2.2.2	Le <i>plain language</i> dans le monde	88
2.2.3	La clarté rédactionnelle en droit dans les pays francophones .	89
2.2.4	Le style clair comme service	90
2.3	Caractéristiques linguistiques du <i>plain language</i>	91
2.3.1	Genres discursifs concernés	92

2.3.2	Synthèse des recommandations rédactionnelles sur le style clair	92
2.3.3	Variabilité des recommandations pour un style clair	100
2.3.4	Exemples de textes estampillés "style clair" avant et après simplification	103
2.3.5	Des textes simples ou simplifiés?	105
2.4	Résumé du Chapitre 2	107
3	Complexité et lisibilité en linguistique	108
3.1	Qu'est-ce que la complexité en linguistique?	109
3.1.1	La complexité, un concept aux multiples facettes	109
3.1.2	Lisibilité, complexité absolue et complexité relative	112
3.1.3	Certains traits linguistiques sont-ils intrinsèquement plus complexes que d'autres?	115
3.2	Approches et mesures de la complexité lexicale	116
3.2.1	Vocabulaire approuvé et non-approuvé	117
3.2.2	La densité terminologique, une mesure du degré de spécialisation?	121
3.3	La complexité dans le groupe verbal	123
3.3.1	Le passif, une source de complexité?	123
3.3.2	Prédicats complexes et procès sémantiques	134
3.4	Complexité et simplification en Traitement Automatique du Langage	138
3.4.1	Mesures automatisées de la complexité lexicale	138
3.4.2	Scores de lisibilité classiques et leurs versions automatiques . .	139
3.4.3	Évaluation automatique de la complexité syntaxique	142
3.5	Résumé du Chapitre 3	148

II	Cadre méthodologique	149
4	Questions de recherche et méthodologie	150
4.1	Différents cadres méthodologiques et théoriques pour étudier le style clair	151
4.1.1	Cadres et réflexions théoriques	151
4.1.2	Approche <i>corpus-based</i> ou approche <i>corpus-driven</i> ?	152
4.1.3	Exploitation des corpus	153
4.1.4	Questions de recherche	155
4.2	<i>Plain language</i> et <i>machine-learning</i> : exploration des métriques de complexité et d'un modèle d'apprentissage pour caractériser les corpus LEX et PLAIN	156
4.2.1	Extraction des métriques de complexité : données et outils	157
4.2.2	Un modèle d'apprentissage-machine pour étudier les corpus LEX et PLAIN	158
4.2.3	Des différences quantitatives entre les genres discursifs du corpus PLAIN?	160
4.3	Approche lexicale et terminologique du style clair en droit	161
4.3.1	Liste de vocabulaire non-approuvé	161
4.3.2	Liste de termes pour calculer la densité terminologique	162
4.3.3	Outils logiciels	164
4.4	Passif et transitivité dans le discours du droit et sa diffusion	165
4.4.1	Le passif dans les corpus LEX et PLAIN	165
4.4.2	Étude de cas : analyse des procès sémantiques avant et après simplification dans des textes qui portent sur la discrimination	167
5	Élaboration et description du corpus	170
5.1	Objectifs et choix du corpus	170

5.1.1	Étudier le discours juridique et sa diffusion dans le cadre du style clair	170
5.1.2	Une approche comparative bilingue français-anglais	171
5.2	Élaborer un corpus : une réflexion sur les genres discursifs	172
5.2.1	Registres et genres du droits choisis	173
5.2.2	Choix des sources discursives : différents acteurs du droit et de la diffusion du droit	174
5.2.3	Le Web comme corpus	175
5.2.4	Ouvrir le domaine spécialisé du droit : fonctions des genres choisis pour le corpus PLAIN	176
5.2.5	Un corpus synchronique ? Espace et temporalité du juridique .	177
5.3	Description des corpus	179
5.3.1	Le Corpus PLAIN	179
5.3.2	Le sous-corpus EN-CAB	179
5.3.3	Le sous-corpus Brochures	181
5.3.4	Le sous-corpus CA-Résumés	182
5.3.5	Les sous-corpus FR-DILA	183
5.3.6	Le Corpus LEX	183
5.3.7	Les sous-corpus Lois	183
5.3.8	Le sous-corpus CA-Jugements	184
5.4	Résumé et outils de préparation des corpus	185
5.4.1	Statistiques	185
5.4.2	Outils de de saisie et préparation des corpus	187
III	Analyse et discussion des résultats	190
6	<i>Plain language et machine-learning</i>	191

6.1	Résultats quantitatifs	192
6.1.1	Performance de la classification des textes LEX et PLAIN à partir des métriques de complexité	192
6.1.2	Quelles sont les métriques de complexité les plus puissantes pour prédire le corpus d'origine d'un texte?	193
6.1.3	Description statistique des métriques les plus prédictives . . .	198
6.1.4	Y-a-t-il des différences entre les genres discursifs du corpus PLAIN?	201
6.2	Flesch Reading Ease et longueur moyenne des unités syntaxiques, des prédicteurs puissants d'après le modèle	204
6.2.1	Le score Flesch Reading Ease, le plus puissant prédicteur pour les corpus LEX et PLAIN	205
6.2.2	Longueur moyenne des phrases et des unités syntaxiques dans les corpus LEX et PLAIN	208
6.3	Analyse qualitative des métriques syntaxiques et lexicales ayant un fort pouvoir de prédiction	215
6.3.1	Pour l'anglais : des métriques qui suggèrent la passage d'un style nominal à un style verbal	215
6.3.2	Pour le français : des différences syntaxiques mais une même complexité lexicale?	218
6.4	Discussion : apports et limites de l'approche quantitative	221
6.4.1	Problèmes méthodologiques dûs aux données	222
6.4.2	Biais et erreurs dans les outils statistiques : peut-on faire confiance aux métriques de complexité?	223
6.5	Résumé du Chapitre 6	227
7	Approche lexicale et terminologique	228
7.1	Les textes anglophones se réclamant du style clair respectent-ils les normes lexicales?	229

7.1.1	Fréquence du vocabulaire "non-approuvé" : résultats quantitatifs	229
7.1.2	Exploration des changements lexicaux par sous-corpus et discussion	231
7.2	La densité terminologique dans les corpus EN-LEX et EN-PLAIN, une mesure du degré de spécialisation?	244
7.2.1	Proportion de termes juridiques : résultats quantitatifs	244
7.2.2	Termes les plus fréquents dans les corpus EN-LEX et EN-PLAIN	246
7.3	Résultats pour le français	251
7.3.1	Mots non-approuvés : quelques exemples	252
7.3.2	Densité des termes dans FR-CA-Jugements et FR-CA-Résumés	254
7.4	Analyse qualitative : Stratégies d'explicitation des mots non-approuvés et des termes spécialisés dans le corpus PLAIN	256
7.4.1	Degrés de spécialisation des termes, termes-clés et dé-terminologisation	257
7.4.2	Définitions, reformulations et utilisation des alternatives claires et simples	258
7.4.3	Quelle influence du genre spécialisé source sur les stratégies explicatives?	264
7.5	Résumé du chapitre 7	267
8	Fréquence et usage du passif dans les corpus LEX et PLAIN	268
8.1	Résultats quantitatifs : fréquence du passif canonique dans les corpus LEX et PLAIN	269
8.1.1	Fréquence du passif canonique dans les corpus anglophones	269
8.1.2	Fréquence des constructions passives canoniques dans les corpus francophones	270
8.1.3	Verbes les plus fréquents dans les passifs canoniques	272
8.2	Formes et fonctions rhétoriques du passif dans les corpus LEX	275

TABLE DES MATIÈRES

8.2.1	Passifs canoniques et <i>boilerplates</i> dans le discours juridique	276
8.2.2	Le passif : un outil pour thématiser les acteurs institutionnels et les objets du droit	279
8.2.3	Passifs sans auxiliaires et cumul de formes passives dans le corpus LEX	286
8.3	Usage et fonctions du passif dans le corpus PLAIN	291
8.3.1	Dans les genres administratifs, passif et thématisation des usa- ger.es du droit et de leur expérience	291
8.3.2	Dans les résumés de jugements, le passif pour thématiser les participant.es de la procédure judiciaire	299
8.3.3	Passifs et fonctions textuelles dans le corpus PLAIN	302
8.3.4	Passifs non-canoniques, cumul de formes passives et voix dis- cursives dans le corpus PLAIN	304
8.4	Résumé du Chapitre 8	307
9	Distribution des procès sémantiques	309
9.1	Résultats quantitatifs : distribution des procès sémantiques dans les échantillons de corpus	311
9.1.1	Résultats pour l'anglais	311
9.1.2	Résultats pour le français	315
9.2	Comparaison des procès sémantiques dans un jugement de la CSC et deux résumés	317
9.2.1	Dans les motifs et résumés en anglais	317
9.2.2	Dans les motifs de jugement et résumés en français	320
9.3	Analyse des procès sémantiques dans les textes de loi sur la discrimi- nation et leur version en style clair	323
9.3.1	Entre EN-CAB et EN-Lois	323
9.3.2	Entre FR-Lois et FR-DILA	326

9.4	Transitivité et modalité dans le discours sur la discrimination : discussion des différences français-anglais	330
9.4.1	Prédicats complexes dans le discours du droit avant et après simplification	330
9.4.2	Modalité et procès sémantiques dans EN-CAB et FR-DILA . .	334
9.5	Résumé du Chapitre 9	336
Conclusion		338
Bibliographie		342
Annexes		359
A	Liste des métriques utilisées pour la classification	360
B	Matrices de confusion obtenues suite à la la classification automatique	365
C	Liste pour l'analyse lexicale et terminologique	368
D	Textes pour l'étude de cas sur les procès sémantiques	373

Table des figures

1.1	Situations discursives du droit (d'après (Maley 1994))	37
1.2	Sous-domaines du droit (Cadou 2014)	39
1.3	Exemple de document en FALC	74
3.1	Mots non-approuvés et leur alternative en anglais	118
3.2	Expressions non-approuvées et leur version claire et simple en français (Commission européenne 2011)	120
3.3	Echelle du Flesch Reading Ease (Flesch 2007)	140
3.4	Métriques de complexité syntaxique du L2SCA (Lu et Ai 2015) me- surées sur un texte législatif et un texte en style clair	145
3.5	Exemple d'utilisation de l'outil Amesure (François et al. 2014)	147
4.1	Relations entre les sous-corpus de EN-LEX et EN-PLAIN	154
4.2	Relations entre les sous-corpus de FR-LEX et FR-PLAIN	154
4.3	Fonctionnement d'un modèle d'apprentissage supervisé	159
4.4	"A to Z" of alternative words - Extrait ©Plain English Campaign	162
4.5	Fonctionnement d'AntWord Profiler (Ferraresi 2019)	164
6.1	Poids des <i>features</i> dans la classification des corpus EN-LEX et EN- PLAIN	194
6.2	Poids des <i>features</i> dans la classification des corpus FR-LEX et FR- PLAIN	196

6.3	Score de Flesch Reading Ease dans les quatre corpus	199
6.4	Scores FRE - Différences entre les genres discursifs du corpus EN-PLAIN	202
6.5	Scores FRE - Différences entre les genres discursifs du corpus FR-PLAIN	203
6.6	MLS et MLT dans les corpus anglophones	209
6.7	MLS et MLC dans les corpus francophones	209
6.8	Énumération dans un texte de FR-LEX	213
6.9	Mise en page d'un texte du corpus FR-PLAIN	214
7.1	Proportion des mots non-approuvés les plus fréquents dans EN-Lois et EN-Brochures	233
7.2	Proportion des mots non-approuvés EN-Lois et EN-CAB	234
7.3	Proportion des mots non-approuvés les plus fréquents dans EN-CA-Jugements et EN-CA-Résumés	239
7.4	Définition des termes dans le discours législatif néo-zélandais (EN-Lois)	265
7.5	Définition des termes au début d'un texte du sous-corpus EN-Brochures	266
8.1	Passifs canoniques et auxiliaires de modalité dans EN-CAB	295
8.2	Passifs canoniques et auxiliaires de modalité dans EN-Brochures . . .	295
8.3	Passifs canoniques avec <i>pouvoir</i> et <i>devoir</i> dans FR-DILA	297
8.4	Passifs canoniques et auxiliaires de modalité dans EN-CA-Résumés .	301
A.1	364
B.1	Matrice de confusion de la classification de EN-LEX et EN-PLAIN .	366
B.2	Matrice de confusion de classification de FR-LEX et FR-PLAIN . . .	367
D.1	Procès sémantiques dans un extrait des motifs en anglais	374

D.2	Procès sémantiques dans un extrait des motifs en français	374
D.3	Procès sémantiques dans un extrait des résumés en anglais	375
D.4	Procès sémantiques dans un extrait des résumés en français	375
D.5	Procès sémantiques dans un extrait de l'Equality Act 2010	376
D.6	Procès sémantiques dans un extrait du Code Pénal français	376
D.7	Procès sémantiques dans un extrait d'EN-CAB sur la discrimination .	377
D.8	Procès sémantiques dans un extrait de FR-DILA sur la discrimination	377

Liste des tableaux

1.1	Exemple de termes juridiques anglophones et leurs définitions (Imprimeur de la Reine pour l'Ontario, 2016)	52
1.2	Extraits du discours juridique anglophone	57
1.3	Extraits du discours juridique francophone	58
1.4	Exemple de reformulation du discours juridique (législatif)	76
2.1	Synthèse des règles rédactionnelles pour un style clair	93
2.2	Extraits de la constitution d'Afrique du Sud avant et après simplification	104
2.3	Gagnant de la catégorie "Best Plain English Sentence Transformation", Plain English Awards, Nouvelle-Zélande, 2018	105
5.1	Le corpus PLAIN	186
5.2	Le corpus LEX	186
6.1	Performance de la classification automatique des textes en se fondant sur les métriques de complexité (en %)	192
6.2	Métriques les plus prédictives pour la classe EN-PLAIN	195
6.3	Métriques les plus puissantes pour prédire la classe EN-LEX	195
6.4	Métriques ayant le plus fort pouvoir de prédiction pour les textes FR-PLAIN	197
6.5	Métriques ayant le plus fort pouvoir de prédiction pour les textes FR-LEX	197

6.6	Score Flesch Reading Ease dans les corpus LEX et PLAIN	198
6.7	Médiane et écart-type des métriques de complexité lexicale et syntaxique les plus prédictives pour l'anglais	200
6.8	Médiane et écart-type des métriques de complexité lexicale et syntaxique les plus prédictives pour le français	201
7.1	Proportion d'items lexicaux "non-approuvés" dans les sous-corpus spécialisés	229
7.2	Proportion d'items lexicaux "non-approuvés" dans les sous-corpus destinés au lectorat non-expert	230
7.3	Proportion de termes dans les différents sous-corpus	244
7.4	Termes de la liste les plus fréquents dans le discours législatif avant et après simplification	246
7.5	Termes de la liste les plus fréquents dans le discours judiciaire anglophone avant et après simplification	247
7.6	Fréquence des mots et expressions "non-approuvées" dans FR-LEX et FR-PLAIN (pmm)	252
7.7	Proportion des termes dans le discours judiciaire et sa simplification (FR-CA) par rapport au nombre total de mots	254
7.8	Termes les plus fréquents dans le discours judiciaire francophone avant et après simplification	255
7.9	Exemple de termes "peu spécialisés" dans EN-PLAIN	257
7.10	Exemple de termes "peu spécialisés" dans FR-PLAIN	258
8.1	Fréquence de la voix passive dans le corpus EN-LEX	269
8.2	Fréquence de la voix passive dans le corpus EN-PLAIN	270
8.3	Fréquence de la voix passive dans le corpus FR-LEX	271
8.4	Fréquence de la voix passive dans le corpus FR-PLAIN	271
8.5	Verbes les plus fréquemment passivés dans le discours juridique anglophone	272

8.6	Verbes les plus fréquents dans les passifs canoniques du corpus FR-LEX273	
8.7	Verbes les plus fréquents dans les passifs canoniques du corpus EN-PLAIN	274
8.8	Verbes les plus fréquents dans les passifs canoniques dans FR-PLAIN	275
9.1	Distribution des types de procès, des modaux et des prédicats complexes dans les textes anglophones (en % du nombre de verbes codés)	311
9.2	Distribution des types de procès, prédicats complexes et auxiliaires modaux dans les échantillons en français (en % du nombre total de verbe)	315

Jetez les yeux sur les détours de la justice. Voyez combien d'appels et de degrés de juridiction ; combien de procédures embarrassantes ; combien d'animaux ravissants par les griffes desquels il vous faudra passer, sergents, procureurs, avocats, greffiers, substituts, rapporteurs, juges, et leurs clerks. Il n'y a pas un de tous ces gens-là, qui pour la moindre chose, ne soit capable de donner un soufflet au meilleur droit du monde. (...) Eh, Monsieur, si vous le pouvez, sauvez-vous de cet enfer-là. C'est être damné dès ce monde, que d'avoir à plaider ; et la seule pensée d'un procès serait capable de me faire fuir jusqu'aux Indes.

Molière, *Les Fourberies de Scapin*,
Acte II, Scène V

Introduction

0.1 Genèse du projet

Le présent travail trouve son origine dans une réflexion commencée dans le cadre de l'accompagnement administratif et juridique de réfugié.e.s et demandeur.euses d'asile en France en 2016, au sein de l'association Solidarité Normalienne de l'ENS Paris-Saclay. Au cours de la procédure de demande d'asile, les demandeur.euses, la plupart du temps allophones, sont confrontés à des textes rédigés par les institutions administratives et juridiques, comme des courriers des tribunaux, dont nous reproduisons un extrait ci-dessous :

Un supplément d'instruction est conféré aux seules fins de permettre à l'Office français de protection des réfugiés et des apatrides, dans un délai de huit jours, de présenter ses observations éventuelles après la réception du mémoire et des pièces complémentaires enregistrées à la Cour. (Ordonnance de la Cour Nationale des Demandeurs d'Asile)

Ce texte contient de nombreux traits caractéristiques du discours du droit considérés comme complexes, notamment l'utilisation du passif, de nombreuses formes nominales et une phrase longue, au ton résolument impersonnel. Devant ce type de textes, c'est un sentiment d'incompréhension, voir d'étrangeté, qui dominait, non seulement chez les demandeur.euses d'asile allophones, mais aussi chez les membres francophones de l'association étudiante. De ce sentiment a découlé une réflexion sur l'accès à l'information juridique pour les usagers du droit. En effet, si ces textes sont difficilement compréhensibles même pour un lectorat francophone, quelles stratégies les institutions mettent-elles en place pour combler ce fossé, et garantir l'accès au droit aux justiciables, dans les diverses procédures et situations qui les concernent ?

Lors de la construction de ce projet de thèse en linguistique anglaise, le choix a alors été fait de travailler sur la communication entre les institutions juridiques et le grand public, anglophones mais également francophones, puisque notre réflexion

avait débuté dans le contexte du droit français. En tant que système qui vise à régir les comportements humains, le droit imprègne toute notre expérience du monde, puisqu'il détermine les comportements autorisés dans nos relations et activités personnelles et professionnelles (pour un.e enseignant.e en classe, par exemple) et crée des relations juridiques entre personnes physiques et morales (par exemple en demandant aux usagers d'adhérer à des conditions générales de vente avant un achat en ligne). Le droit est par ailleurs un domaine profondément linguistique : c'est essentiellement par le langage que ces relations sont créées (Gibbons 1999). Or, si c'est un lieu commun de dire que la loi doit être connue de tous.tes, la complexité du système judiciaire et de sa langue est également proverbiale, et ce depuis bien longtemps¹. En France comme dans les pays anglo-saxons, des voix s'élèvent régulièrement pour rappeler que la bonne compréhension du discours juridique constitue un enjeu politique et social majeur. Cette compréhension conditionne l'exercice des droits civiques et politiques² et l'accomplissement de démarches administratives permettant l'accès aux droits sociaux et au soins dans un contexte de numérisation croissante. Linguistiquement, ces questionnements se traduisent par les questions suivantes : Les textes que les institutions publient à l'attention des justiciables sont-ils véritablement lisibles ? Comment passe-t-on de textes spécialisés perçus comme complexes à des textes compréhensibles par des non-spécialistes ?

Dans les pays anglophones, depuis plusieurs décennies, on a tenté de répondre à ces questions grâce au Plain Language Movement, qui a promu le concept de *plain language*, défini comme un style linguistique qui permet une communication efficace avec les usagers (Petelin 2010). Le *plain language* repose sur une série de recommandations qui incluent l'utilisation d'un vocabulaire courant, de phrases courtes et de la voix active Rossetti et al. (2020). Ces règles pour rédiger en style clair ont été adoptées par certaines institutions juridiques, en particulier dans le cadre de la communication avec le grand public. C'est le cas notamment au Royaume-Uni et en Nouvelle-Zélande, pays pionniers du Plain Language Movement, mais aussi au Canada, où le concept a été adapté à la politique de bilinguisme et donc au français (Asprey 2004). Ces pays ont donc été choisis pour constituer nos corpus. En France, bien que le concept de *plain language* n'existe pas en tant que tel, la langue juridique, en particulier administrative, a fait l'objet d'efforts de simplification (Collette et al. 2002), qui se fondent sur le même type de règles que le *plain language*, par exemple éviter le passif et utiliser des mots familiers. C'est pour faire référence à ces efforts

1. Comme le suggère la citation de Molière présentée en épigraphe.

2. La lisibilité des bulletins de vote aux Etats-Unis est une question récurrente dans la presse, par exemple en 2016 : <https://www.nytimes.com/roomfordebate/2016/11/03/why-are-ballot-measures-so-confusing/ballot-measures-need-to-be-written-in-plain-language>

de simplification dans ces différents contextes, que nous employons, dans le titre et dans la suite de notre travail, l'expression "style clair" pour désigner les efforts de simplification de la langue. Le terme *style* est ici entendu comme un ensemble de conventions linguistiques acceptées (Commission Européenne 2020). Ce terme nous permet ainsi d'inclure les efforts de la clarté et la simplicité dans la rédaction mais qui ne se rattachent pas strictement au *Plain Language Movement*, par exemple en France.

0.2 Problématique et objectifs

Ce travail a pour but d'étudier la reformulation du discours du droit dans le cadre de la communication entre les institutions et les justiciables, en français et en anglais, à l'aune des recommandations pour un style clair. Or, ces recommandations s'appuient elles-mêmes sur certaines approches de ce que constitue la complexité en linguistique, par opposition à la clarté et à la lisibilité. Notre objectif est donc également d'interroger cette vision de la complexité et de confronter les outils conceptuels et méthodologiques qui ont été développés pour la mesurer à des productions textuelles réelles.

Pour cela, nous avons constitué deux corpus bilingues, l'un intitulé LEX, le second PLAIN. Le premier est constitué de textes destinés à un public spécialisé, le second à un lectorat non-spécialisé. Le discours du droit recouvrant une diversité de situations de communication et de types de textes, ils sont composés de plusieurs sous-corpus qui correspondent à différents genres discursifs et différentes fonctions. Concernant le domaine du droit, le corpus LEX est composé de textes qui représentent le discours législatif et le discours judiciaire. Pour le corpus destiné au grand public, les textes incluent des brochures et pages web destinées à guider les justiciables dans les procédures juridiques, ainsi que des résumés de décisions de la Cour Suprême du Canada.

Au vu de la diversité du domaine du droit et de la notion de style clair, nous avons mobilisé plusieurs approches théoriques et méthodologiques. Notre travail s'appuie en premier lieu sur les études en anglais de spécialité et en analyse de discours. Nous adoptons, à leur suite, des approches et méthodes classiques de la linguistique de corpus, en nous intéressant notamment à la terminologie et la phraséologie du droit et à certains traits linguistiques précis (dans une approche *corpus-based*). Afin d'éprouver l'outillage de l'analyse de discours et de la complexité linguistique, nous

avons également choisi des méthodologies issues du Traitement Automatique des Langues et de l'acquisition de langue seconde pour explorer le passage du corpus LEX et au corpus PLAIN. Enfin, notre approche des textes est résolument fonctionnelle, et nous portons une attention particulière aux notions de genres discursifs et de fonctions rhétoriques, en empruntant notamment certains outils de la Grammaire Systémique Fonctionnelle. Toutes ces approches diverses nous ont semblé nécessaires pour éclairer d'un jour nouveau les phénomènes discursifs qui nous occupent.

Cette thèse est organisée autour de trois objectifs principaux :

1. Confronter les recommandations pour un style clair et la vision de la complexité qui les sous-tend aux productions réelles des institutions juridiques.
2. Proposer une méthodologie systématique, utilisant des outils à la fois classiques et innovants, pour étudier le passage d'un discours réputé complexe à un discours de médiation juridique.
3. Comparer les stratégies de reformulation du discours du droit selon la fonction des textes simplifiés et le contexte culturel (notamment la langue).

L'organisation de cette thèse résulte de ces objectifs et des approches adoptées. Dans la première partie, nous proposons une revue de la littérature afin de définir les problématiques et approches théoriques dont relève notre travail. Nous nous attachons d'abord à dépeindre le discours du droit et de sa diffusion auprès du grand public dans leur diversité, en portant une attention particulière aux diverses situations de communications et genres textuels qu'ils recouvrent, ainsi qu'à leur caractéristiques linguistiques et discursives (Chapitre 1). Nous présentons ensuite l'histoire et les traits linguistiques du style clair, en proposant une synthèse des recommandations appliquées par plusieurs institutions à travers le monde (Chapitre 2). Dans le Chapitre 3, nous discutons des notions de complexité linguistique et de lisibilité sur lesquelles reposent le style clair, mais également des outils et mesures qui ont été développés pour les mesurer grâce à la recherche en acquisition du langage et en traitement automatique des langues.

Dans la seconde partie, nous délimitons le cadre méthodologique de notre travail. Nous présentons d'abord les approches et outils que nous avons adoptés (Chapitre 4). Nous décrivons ensuite les corpus que nous avons constitués (Chapitre 5).

La troisième partie est consacrée à la présentation et à l'analyse des résultats. Nous proposons d'abord une approche quantitative et statistique du style clair : une méthode de *machine-learning* est utilisée pour classer automatiquement les textes des corpus selon leur degré de spécialisation (LEX ou PLAIN), en nous appuyant sur des métriques de complexité développées en TAL en en acquisition du langage (Chapitre 6). Le Chapitre 7 adopte une approche lexicale et terminologique du style clair en droit, tandis que le Chapitre 8 présente une analyse des formes de passif dans nos corpus. Enfin, dans le Chapitre 9, nous rapportons une étude de cas portant sur le concept juridique de discrimination au prisme des procès sémantiques, avant de proposer une conclusion générale.

Première partie

Cadre théorique

Chapitre 1

Caractérisation du discours juridique et de sa diffusion auprès des non-spécialistes

Nous présentons dans ce chapitre une revue de la littérature portant sur le discours du droit. Avant de présenter la notion de style clair, il est en effet nécessaire de proposer une caractérisation linguistique et discursive du domaine du droit. L'objectif est ainsi de décrire les modes et types de texte sur lesquels porte notre travail, afin de mieux saisir les enjeux du *plain language* dans ce domaine, en particulier dans la communication des institutions juridiques avec les non-spécialistes. Dans un premier temps, nous nous attachons donc à décrire les différents contextes, situations de communication, participants et caractéristiques linguistiques qui forment le registre du droit. Dans un second temps, nous présentons les relations de ce domaine avec le grand public, ainsi que les enjeux et traits discursifs de sa diffusion (*dissemination*) auprès d'un lectorat non-expert.

1.1 Le domaine du droit et son étude linguistique

Nous présentons d'abord le domaine du droit, afin de contextualiser l'étude du registre juridique au sein de ce domaine spécialisé. Un domaine spécialisé est entendu comme "[l']ensemble de connaissances et/ou de pratiques mis au service d'une même finalité" (Yeught 2016). Comme nous allons le voir, ce domaine renvoie à une multiplicité de contextes et de professions. Ces différents acteurs agissent pour accomplir

les missions du droit, qui correspondent à des fonctions "de direction des conduites humaines et de gestion des conflits" (Deumier 2019 : 19), incarnées notamment par la règle de droit et la figure du juge.

1.1.1 La langue et le droit

Il faut en premier lieu souligner le lien fondamental qui existe entre le droit et la langue. Les langues de spécialité de la médecine, des mathématiques ou encore de la sociologie, visent toutes à décrire des objets, événements et relations qui ont une réalité extra-linguistique, réalité qu'ils conserveraient même si ces discours n'existaient pas. Dans le cas du droit, bien que celui-ci vienne réguler des relations sociales extra-linguistiques, c'est par la langue que se créent les relations de droit. Gibbons définit le droit comme "la plus linguistique des institutions" (Gibbons 1995), et Maley rappelle que le droit est intrinsèquement lié à la langue :

Language is **medium, process and product** in the various arenas of the law where legal texts, spoken or written, are generated in the service of regulating social behavior. (Maley 1994 : 11)

Le droit, dont la finalité est de régir les comportements et relations humaines par la règle, le fait en effet *à travers* le discours. Plus encore, c'est la dimension performative du discours juridique qui crée ce lien essentiel entre la langue et le droit, comme le décrit Stroup :

[Law] is language – not only language, but a very special kind of language, for law is an attempt to structure the realities of human behavior through the use of words. When a legislature passes a law or a court hands down a decision, it is altering the status of individuals, changing their relationships to other individuals, to possessions and objects, to the state. Legal language does not merely describe these relationships ; it affects what it describes. (Stroup 1983)

C'est donc par la production linguistique que les institutions juridiques et administratives font advenir, ou au contraire annulent, les relations entre les personnes et les biens, entre les humains et les institutions. Cette exacerbation de la dimension linguistique imprègne le droit à tous les niveaux, que ce soit dans les amendements d'un texte législatif qui vient inscrire une nouvelle formulation dans la loi, ou, lors d'un procès, lorsque les avocat.e.s et juges reviennent sur les faits racontés par les témoins et accusés. Il n'est donc pas étonnant que juristes, linguistes, et analystes du discours aient produit une abondante littérature qui explore les liens entre langue, discours et droit, selon différentes approches et dans des contextes culturels divers.

1.1.2 Familles juridiques et territorialité du droit

Le droit est en effet, comme la langue, un produit socioculturel et culturellement spécifique, inscrit dans une réalité historique dont on ne peut le détacher (Deumier 2019 : 166) ; il est intrinsèquement lié à "une dimension spatiale et temporelle" (Monjean-Decaudin 2010) puisqu'il s'applique a priori à un moment donné, sur un territoire donné. La dimension temporelle sera évoquée plus loin (voir Chapitre 5), mais il est important de discuter dès à présent de la dimension territoriale du droit. Le droit et les systèmes juridiques dans lesquels ils s'inscrivent sont le résultat d'un long processus historique. On distinguera deux grandes familles juridiques d'où sont issus les textes anglophones et francophones étudiés.

- La famille de *Common law*, originaire du Royaume-Uni, et plus particulièrement d'Angleterre, qui a essaimé dans les pays du Commonwealth ;
- La famille de droit civil, dont le droit français est l'archétype et le modèle dominant (Deumier 2005 : 167).

Historiquement construit autour de la figure du juge, le système de *common law* a pour source déterminante du droit le précédent judiciaire et est centré autour des actions en justice, là où le système de droit civil réserve à la loi une place prépondérante, symbolisée par le Code civil, "véritable marque de fabrique de la famille" et symbole de "l'idéal d'un droit complet" (Deumier 2019).

Les productions textuelles anglophones étudiées dans ce travail proviennent du Royaume-Uni, de la Nouvelle-Zélande et du Canada. Ces États ont été choisis pour leur politique linguistique d'utilisation du style clair dans la communication du droit au grand public, comme nous le verrons plus bas. Cependant, ces textes seront à la fois des exemples de *case law*, puisque nous étudierons des textes de la Cour Suprême du Canada, mais aussi des textes législatifs faisant partie de la *statute law*, dont la symbolique dans le raisonnement des juges est différente de celle du droit civil (Deumier 2005 : 171). Au Canada, la mixité des familles juridiques est illustrée par un système de bijuridisme. A cause de l'histoire coloniale du pays, se côtoient un système de *common law*, mis en place dans les anciennes colonies britanniques, et le droit civil puisque le Québec, ancienne colonie française, en est l'héritier. Ce système de bijuridisme se double d'une politique bilingue : les textes francophones de la Cour Suprême du Canada ont donc également été choisis pour notre corpus. Les productions des instances législatives et administratives issues du droit de France semblent en effet constituer point de comparaison avec celles des pays anglophones ou bilingues cités plus haut, qui ont depuis des décennies établi une politique linguistique d'utilisation du style clair.

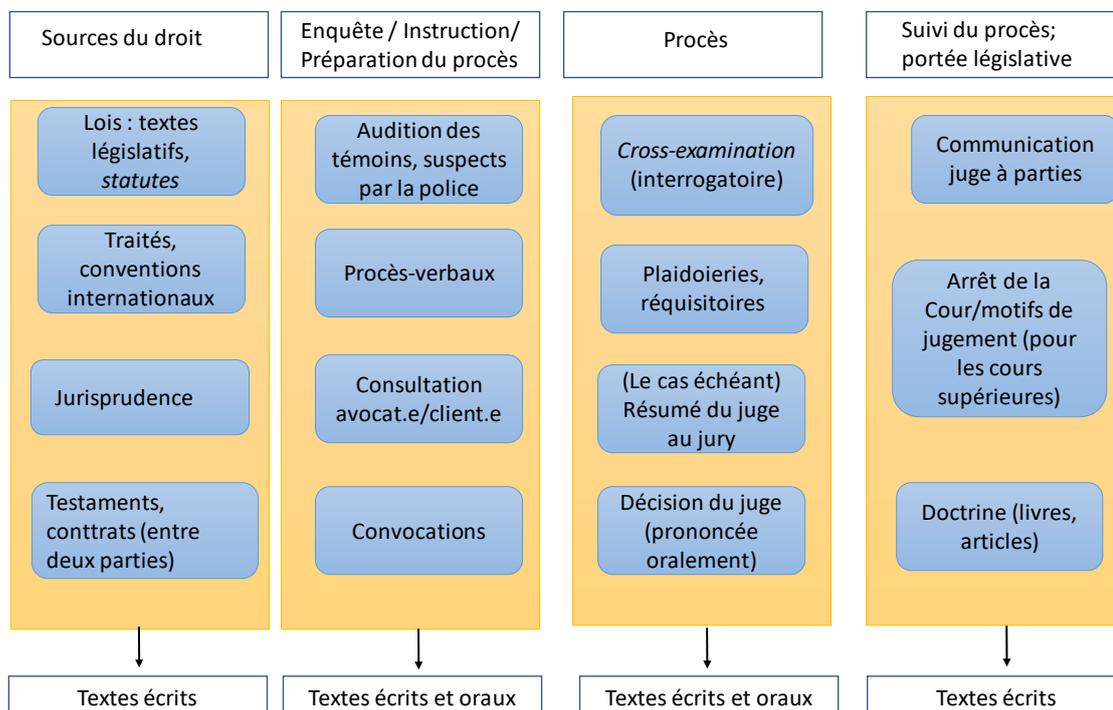
Le français juridique utilisé au Canada n'est cependant pas le même que celui utilisé en France, puisque les deux systèmes juridiques n'appartiennent pas à la même famille. Cette divergence appelle d'ores et déjà à une grande prudence méthodologique, en particulier concernant la terminologie des deux systèmes et des procédures qu'ils renferment. Pour la partie anglophone de notre étude, la prudence est également de mise. Cependant la Nouvelle-Zélande et le Canada font partie tous deux du Commonwealth ; leur système est donc héritier, et culturellement proche de celui du Royaume-Uni : les termes utilisés, par exemple pour décrire le système des cours d'appel, est donc le même, à peu d'exceptions près. Il faut malgré tout garder à l'esprit les spécificités juridiques de ces différents États anglophones. En particulier, le droit auquel nous nous intéresserons pour le Royaume-Uni sera le droit anglais, non pas par exemple le droit écossais, qui comporte certaines différences (par exemple pour les questions de succession ou la séparation des biens suivant un divorce) ; ce choix a été fait par mesure de simplicité.

1.1.3 Le droit : un domaine spécialisé hétérogène

Nous avons jusqu'ici évoqué le droit comme un domaine spécialisé uniforme, mais il est caractérisé par une grande hétérogénéité de pratiques et situations discursives. Ce domaine spécialisé renvoie en effet à de nombreux contextes de communication entre les membres de différentes communautés de discours. Le domaine du droit, puisqu'il vient régir les relations interpersonnelles et institutionnelles des justiciables, fait par ailleurs partie, comme la médecine, des domaines spécialisés qui "par leur finalité même, [sont] en contact étroit avec le grand public et les interfaces sociale et linguistique entre le spécialisé et le non-spécialisé sont nombreuses et actives." (Yeught 2016).

Ainsi les justiciables sont-ils en contact avec les spécialistes du domaine dans une variété de situations discursives présentées dans la Figure 1.1, qui présente les situations de discours juridiques en contexte anglo-saxon inspirées par Maley (1994). Quatre situations de discours principales apparaissent : les situations sources du droit, symbolisées en particulier par la création de textes législatifs et de régulations par les autorités parlementaires et gouvernementales, mais également les précédents dans la tradition de *common law* (voir Section suivante) ; la procédure judiciaire avant un procès, où interagissent la police avec les témoins, victimes et suspects d'une affaire, mais également ces dernier.e.s avec leurs avocat.e.s ; le procès, où les juges, avocat.e.s et parties interagissent les uns avec les autres ; enfin, les situations

FIGURE 1.1 – Situations discursives du droit (d’après (Maley 1994))



après un procès où les juges notamment communiquent leurs décisions. On peut distinguer deux modes de discours fondamentaux du droit : l’oral et l’écrit, le premier étant caractérisé par un degré de formalité moindre (comme c’est le cas pour les usages non-spécialisés de la langue) (Tiersma, 1999 : 135). Ces modes et situations discursives diverses donnent lieu à des productions textuelles nombreuses, et donc à divers de types de textes, appelés genres discursifs. Un genre discursif peut être défini comme un événement communicatif structuré, délimité par une situation socio-historique, un contexte de production ainsi que par des objectifs communicatifs qui se traduisent par des contraintes linguistiques et textuelles :

A genre, here, is defined as a category assigned on the basis of external criteria such as intended audience, purpose, and activity type. That is, it refers to a conventional, culturally recognised grouping of texts based on properties that configure the internal (linguistic) criteria forming the basis of text type categories. (Orts Llopis 2009)

Nous reviendrons plus longuement sur cette définition, puisque la suite de ce travail adopte une approche *genre based* du discours du droit et de sa diffusion, mais on retiendra ici que les situations discursives évoquées ci-dessus entraînent donc la production de nombreux genres textuels typifiés, par exemple les procès-verbaux de la police, les notifications et avis envoyés par les gouvernements et institutions

judiciaires aux justiciables, ou encore les contrats, testaments, textes de lois - tout un ensemble de textes que Bhatia nomme une **constellation de genres discursifs** (Bhatia 2004) parce qu'ils sont reliés entre eux. Ainsi les genres discursifs oraux incluent-ils par exemple les auditions de témoins et suspect.e.s, lors d'une enquête de police, qui sont ensuite transcrits sous forme de procès verbaux écrits, eux-mêmes utilisés par les avocat.e.s et magistrat.e.s au cours des procès, notamment dans le cadre des interrogations de témoins ; les genres écrits sont par exemples les textes législatifs, ou encore les décisions de justice écrites, qui s'appuient à la fois sur les genres oraux de la procédure judiciaire et les textes de loi produits par les législateur.ices pour rendre une décision, oralement puis par écrit. L'interconnexion des textes juridiques (Office of the Scottish Parliamentary Counsel 2006), produits dans des situations discursives différentes, est liée à la variété de locuteurs et interlocuteurs du droit, professionnels ou profanes, qui fait dire à Cornu : "Le droit a mille bouches", d'où sortent des "voix qui se mêlent dans la création et la réalisation du droit" (Cornu 2005 : 214).

On pourrait ajouter que le droit a également mille branches, car il est par ailleurs composé de nombreux sous-domaines spécifiques, comme le montre la Figure 1.2. Le droit est souvent assimilé à un arbre aux ramifications complexes, et on voit sur l'image que le droit constitutionnel, droit fondateur d'un pays, n'est pas sur la même "branche" que le droit administratif ou que le droit international. Les professions liées au droit sont donc elles-mêmes spécialisées : c'est pourquoi il existe par exemple des juges et tribunaux administratifs, qui ne jugent pas le même type d'affaires que les cours pénales. Une affaire ne sera pas jugée à la même cour selon les circonstances des faits. Par exemple, les patients d'un hôpital public verront leurs litiges tranchés selon les règles du droit administratif, tandis que ceux d'une clinique privée, seront soumis aux règles du droit privé (Cadou 2014). On voit donc apparaître ici une forme de complexité structurelle et d'hyperspécialisation du domaine du droit, qui peut être difficile à naviguer pour les justiciables non spécialistes.

L'étude d'un discours juridique demande donc que l'on se situe vis-à-vis du domaine, qui est, comme nous l'avons vu, hétérogène (Gautier 2003). Dans le cas présent, nous avons choisi de nous concentrer sur le droit national, en excluant donc le droit communautaire européen (pour les pays de l'Union Européenne comme la France¹) ainsi que le droit international. Concernant le droit interne, nous avons exclu certaines branches comme le droit fiscal, le droit des finances publiques, le droit commercial. Les sous-domaines du droit abordés sont variés et incluent no-

1. Et, à l'époque où notre corpus a été constitué, le Royaume-Uni.

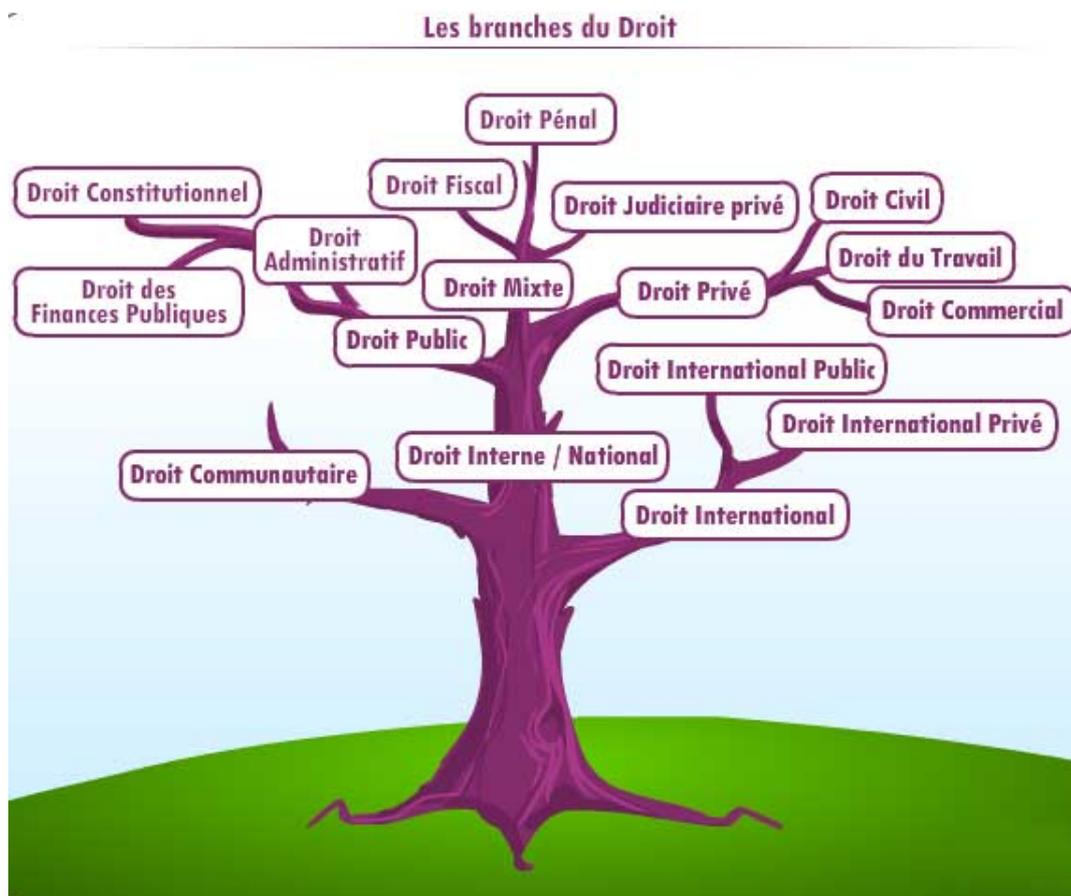


FIGURE 1.2 – Sous-domaines du droit (Cadou 2014)

tamment le droit pénal, le droit du travail et le droit administratif (en particulier droit des étrangers), puisque les situations qui nous intéressent sont les moments de contact entre justiciables et institutions juridiques, dans des circonstances aussi diverses qu'une agression, une arrestation, la perte d'emploi liée à une discrimination ou l'adoption d'un enfant.

1.1.4 Différentes approches des discours du droit

Au vu de cette diversité du domaine du droit, la recherche sur le discours juridique a adopté plusieurs approches. On peut souligner que la traductologie est au premier plan de l'étude des discours juridiques, dans une ère d'internationalisation et de déterritorialisation globale du droit, qui pose notamment des questions terminologiques pour la traduction de concepts culturellement spécifiques. Par exemple, la traduction de documents juridiques dans le cadre d'institutions multilingues nationales ou internationales, comme dans l'Union Européenne a attiré l'attention des

chercheur.e.s (Biel 2004), de même que ses applications pour l'enseignement de la traduction juridique (Orlando 2018).

Une deuxième approche majeure du discours juridique est celle liée au cadre de la procédure judiciaire, et à la langue comme matériau pouvant servir de preuve au tribunal ; ses méthodes puisent à la fois dans la linguistique de corpus et l'analyse de discours. Il s'agit de la linguistique forensique ou légale (*forensic linguistics*), définie ainsi :

Forensic linguistics is concerned with the provision of expert linguistic evidence, usually in court. (Gibbons, 1994 : 319)

La linguistique forensique est donc liée au rôle des linguistes comme expert.e.s., dans des cas où les preuves récoltées ont pour matériau la langue, qu'elle soit orale (dans le cadre) ou écrite (dans des affaires de plagiat notamment). Plusieurs genres discursifs issus de la procédure judiciaire ont été étudiés, notamment les appels aux services d'urgence ou les auditions de témoin, (Coulthard et al. 2016 : 71-94), ou encore les interactions lors d'un procès, considérées comme formant un genre discursif complexe (Coulthard et al. 2016 : 96). Les textes de la procédure judiciaire étant souvent personnels et confidentiels, et leur récolte posant donc des questions d'anonymisation, ils ont été exclus de notre étude. Nous avons exclu également le discours du tribunal (*courtroom discourse*) pour nous concentrer sur la communication entre institutions juridiques et grand public, mais il existe de nombreuses études linguistiques et sociolinguistiques portant sur les interactions lors des procès, par exemple entre victimes et avocat.e.s (Matoesian 1997) ou des échanges entre magistrats.

Notre travail s'inscrit quant à lui dans la continuité des études, nombreuses, d'analyse de discours et de linguistique de corpus portant sur l'anglais juridique. Les termes de *legal linguistics* ou de jurilinguistique sont utilisés dans différents contextes culturels pour parler de l'étude du discours spécialisé juridique, c'est-à-dire des usages spécialisés de la langue dans les contextes juridiques. En particulier, l'approche que nous adoptons est celle du domaine de l'anglais de spécialité (et du français de spécialité). Cette branche de l'anglistique, en anglais *English for Specific Purposes*, est la branche de l'anglistique qui s'intéresse aux usages spécialisés de la langue, souvent dans une perspective appliquée à l'enseignement ; il est équivalent dans ce cadre au Français sur Objectifs Spécifiques (Lehmann 1997). Plus largement, il s'agit du domaine qui s'intéresse à l'étude des langues et discours spécialisés : fondamentalement pluridisciplinaire, il emprunte à des méthodologies

et positionnements théoriques aussi divers que la terminologie et la lexicographie, l'analyse de discours, la linguistique de corpus, ou encore l'ethnographie. Ce domaine se fonde sur la notion de communautés de discours spécialisées (elle-même reposant sur la notion de domaines spécialisés)², c'est-à-dire l'ensemble des personnes qui œuvrent aux finalités d'un même domaine (Yeught 2016), dans notre cas, le droit. Yeught (2016) précise que les communautés de discours peuvent être restreintes à un territoire national :

La communauté spécialisée anglophone considérée comporte plusieurs communautés spécialisées nationales qui œuvrent à la finalité du domaine dans un cadre national donné : Grande-Bretagne, États-Unis, Canada, Irlande, Australie, Nouvelle-Zélande, etc. (Yeught 2016)

Cette précision est importante en particulier pour le droit, puisque, comme mentionné plus haut, notre travail porte sur différents pays anglophones et francophones, où les systèmes juridiques diffèrent parfois, bien qu'ils appartiennent respectivement aux familles de *common law* et droit civil. Dans notre travail, c'est donc la relation entre cette communauté de discours spécialisée, qui est en fait composée de plusieurs milieux professionnels, et le public non spécialisé (donc extérieur à la communauté de discours) qui nous intéresse. Avant de décrire cette relation, il convient de circonscrire les types de discours et de textes juridiques sur lesquels portent ce travail.

1.2 Typologie des discours du droit étudiés

Sauf indication contraire, tous les extraits et exemples dans la suite de ce chapitre sont tirés de nos corpus de thèse. Ce corpus est constitué de deux corpus principaux : un corpus de textes juridiques non simplifiés, LEX, c'est-à-dire destiné principalement à un lectorat expert ; et un corpus rédigé en style clair, PLAIN, destiné principalement à un lectorat non spécialiste. Nous décrivons en détail ces corpus au Chapitre 4, mais les utilisons dans ce chapitre comme source d'exemples.

1.2.1 Le registre juridique

Comme nous l'avons évoqué, le discours du droit est pluriel, en termes de branches et de sous-domaines, mais également en termes de contextes de produc-

2. Ainsi que des milieux et pratiques spécialisées.

tion (au commissariat, au tribunal, au parlement, dans les administrations, etc.). Il n'existe donc pas en réalité *un* mais *des* discours du droit, une variété qui découle elle-même de la diversité de situations englobées sous l'adjectif "juridique" :

There is not one legal discourse but **a set of related legal discourses**. Each has a characteristic flavour but each differs according to the situation in which it is used. (Maley 1994)

La variété linguistique juridique, reflétant la pluralité du domaine mais unifiée par un style caractéristique et des objets communs, peut donc être qualifiée de **registre** : c'est une variété linguistique utilisée pour des objectifs et dans des situations de communication spécifiques :

Legal language is a type of register, that is, a **variety of language appropriate to different occasions and situations of use**, and in this case, **a variety of language appropriate to the legal situations of use**. (Cao 2007 : 9)

Cette variété interne invite donc à l'envisager dans une perspective de registre (*register perspective*), c'est-à-dire à étudier les types de textes que renferme ce registre selon leur contexte socio-pragmatique d'utilisation, de production et selon leurs fonctions de communication (Conrad et Biber 2005; Pic et Furmaniak 2014). En ce qui concerne le droit, la variété des situations de production a entraîné l'apparition de différents sous-registres, comme nous allons le voir.

Cao (2007) inclut ainsi dans le registre juridique non seulement le langage de la loi (*language of the law*), mais également le langage portant sur la loi (*language about the law*), qui est donc au cœur de notre travail, et le langage utilisé dans d'autres contextes de communication juridique (par exemple, la correspondance entre deux avocats). L'auteur distingue alors quatre sous-registres juridiques, que nous avons plus haut appelés discours, et qui sont eux-mêmes divisés en genres discursifs ou types de textes différents, aux statuts juridiques différents et classés selon leurs fonctions communicatives :

- les textes législatifs, produits par les autorités législatives et qui incluent les propositions et textes de loi, les traités internationaux et les lois plurilingues ;
- les textes judiciaires, produits au cours de procédures judiciaires par les agents de la justice et autres autorités légales (préfectures, etc.) ;
- les textes juridiques académiques produits par des juristes (universitaires) ou des avocats, désignés dans le système juridique français par le terme "doctrine" ;
- les textes juridiques privés (*private legal texts*), qui incluent selon Cao (2007) les documents écrits par les avocats et juristes, dont notamment les contrats

privés, baux, testaments, les documents relatifs aux litiges, mais également des documents produits par des non-avocats, comme des accords privés ou dépositions de témoins.

Les textes de la doctrine n'ont pas été retenus comme un registre à représenter dans notre corpus, car ils correspondent avant tout à un registre académique. Bien que la dissémination de la recherche soit également un enjeu du style clair, nous avons préféré les exclure de notre corpus car ils sont le fruit d'une communication très spécifique entre expert.es du droit. Les textes du dernier sous-registre (*private legal texts*) ont d'abord été envisagés pour constituer notre corpus, mais ils ont également été exclus. Ils consistent généralement en des documents personnels et confidentiels relevant d'une correspondance privée entre avocats, institutions judiciaires et justiciables ; ils sont donc difficiles à récolter, même sous la forme d'échantillons, pour des raisons de respect de la confidentialité. Les sous-registres du droit qui nous intéressent sont donc les registres législatif et judiciaire.

1.2.2 Le discours législatif

Considéré comme le plus ésotérique des discours juridiques, le discours législatif est aussi celui auquel les justiciables sont le plus fréquemment exposés (Maley 1994). Suivant la définition de Richard (2006), les textes du discours législatif ont force de loi, c'est-à-dire qu'ils sont formés d'énoncés prescriptifs qui incluent les constitutions, les lois et propositions de lois, les conventions et régulations internationales. Au même titre que les autres documents normatifs (Richard 2006) que sont les contrats ou les testaments, les textes législatifs ont une fonction performative. Ils sont par excellence des *speech acts* (Searle 1965) : par le simple fait de l'énoncer, ils réalisent la promulgation de la loi, et créent ou modifient des relations juridiques (c'est d'ailleurs pour cette raison qu'on appelle les textes de loi *Acts of Law* en anglais).

Nous nous intéressons au genre des textes législatifs issus du droit national, en anglais, *statutes*, en français, textes de lois et décrets ou arrêtés du gouvernement. Ces textes normatifs énonçant la règle juridique sont produits par les parlements et les gouvernements, et c'est sur eux que s'appuient les procédures judiciaires et administratives auxquelles sont confrontées les justiciables et citoyen.nes. Bien qu'ils soient publics, ils visent un lectorat spécialisé dans le domaine juridique, formé principalement par les avocat.e.s et juges qui sont chargés d'interpréter et d'appliquer la loi. Leur caractère performatif est particulièrement visible dans les formules rituelles qui encadrent et parsèment les textes de loi.

- (1) Be it enacted by the Queen's most Excellent Majesty, by and with the advice and consent of the Lords Spiritual and Temporal, and Commons, in this present Parliament assembled, and by the authority of the same, as follows... (Formule d'ouverture d'un Act du parlement britannique)

La formule figée et subjonctive *be it enacted* dans l'exemple ci-dessus illustre à la fois l'acte performatif et le caractère solennel de la loi, qui ici se veut adressée à tous. On touche ici au paradoxe du texte législatif, qui a pour usagers principaux des juges, avocats et semi-spécialistes (agents administratifs, travailleurs sociaux), mais doit également s'adresser au plus grand nombre puisque, le dit l'adage (souvent cité par les juristes comme Deumier (2019) "Nul n'est censé ignorer la loi"). Ainsi Maley (1994) souligne-t-il le rôle difficile des rédacteur.ices des textes législatifs, qui doivent s'assurer que la loi est assez précise et non ambiguë afin d'être interprétée par les avocat.e.s et les juges selon l'intention originelle des législateurs, tout en s'appliquant et en s'adressant aux citoyen.ne.s. Depuis quelques dizaines d'années, l'accessibilité accrue des textes législatifs grâce à la numérisation a également modifié le profil du lectorat des textes de lois. Les résultats de l'enquête *Good Initiative Survey* au Royaume-Uni, rapportés par Xanthaki (2019), offrent un portrait contrasté des usagers des textes de loi. La majorité des utilisateurs du site *legislation.gov.uk* (60%), d'après une étude sur deux millions de visiteurs, est en fait constituée de professionnels non spécialistes du droit dont l'activité est fondée sur la loi et les textes législatifs (plus précisément, policiers, agent.e.s des autorités locales, professionnel.les des ressources humaines). Dans les 40% restants, la moitié se compose de non-spécialistes (*laypersons*) qui consultent ces textes afin de voir comment la loi s'applique à leur propre situation ou à celle de leurs proches. Il apparaît donc que seulement 20% des personnes qui consultent les textes de lois sont des professionnel.les du droit (avocat.e.s, juges, juristes) :

There is now compelling empirical evidence that legislation, at least in the UK, is mainly read by lay users, namely users without legal sophistication who read legislation in order to acquire answers to problems related to their everyday life or to the exercise of their professional duty. (Xanthaki 2019)

C'est pourquoi le *Plain Language Movement* a donc eu pour objet non seulement la communication entre les institutions et les justiciables, mais également la rédaction de la législation elle-même dans certains pays comme le Royaume-Uni, la Nouvelle-Zélande et l'Australie (Williams (2015)), le style clair étant alors perçu comme politiquement nécessaire pour exprimer clairement la loi (Office of the Scottish Parliamentary Counsel 2006). C'est pourquoi nous avons choisi le Royaume-Uni

et la Nouvelle-Zélande comme source de textes législatifs, où ont également été implémentées les recommandations du PLM pour la rédaction de la législation.

Le genre législatif est caractérisé par la division en chapitres, sections, articles (en français), sous-sections (en anglais) et alinéas, et par des renvois fréquents à des articles ou chapitres à l'intérieur d'un même texte législatif, ou à d'autres textes de lois (*cross-referencing*). On peut dégager plusieurs fonctions rhétoriques récurrentes à l'intérieur ces différentes sous-divisions, illustrées par les exemples ci-dessous.

- (2) Where a divorce takes effect in accordance with this section, **a judge or officer of the court** (...) or, where that judgment has been appealed, of the appellate court that rendered the judgment on the final appeal, **shall, on request, issue to any person a certificate that a divorce granted under this Act dissolved the marriage** (...) effective as of a specified date. (EN-Lois)
- (3) Lorsque le divorce est prononcé, si des dettes ou sûretés ont été consenties par les époux, solidairement ou séparément, dans le cadre de la gestion d'une entreprise, **le tribunal de grande instance peut décider d'en faire supporter la charge exclusive au conjoint qui conserve le patrimoine professionnel** (...) (FR-Lois)

Selon la terminologie employée par Bhatia (1994) trois types de dispositions peuvent être distinguées. Tout d'abord, on y trouve les *Preparatory qualifications*, les dispositions qui décrivent les cas et situations auxquelles s'applique la loi en question, introduites par des propositions conditionnelles (souligné dans les exemples). Viennent ensuite les *Operational qualifications*, qui donnent des indications sur l'exécution et/ou le fonctionnement de la loi (en gras dans les exemples), pour expliquer le rôle de la juridiction ou administration qui sera chargée d'appliquer la loi, par exemple dans l'exemple 3 *le tribunal de grande instance peut décider, etc.* Enfin, à divers endroits, on trouve les *Referential provisions*, c'est-à-dire toute référence à d'autres textes législatifs ou d'autres passages du même texte, ce qui rappelle l'interconnexion des textes de loi et son caractère intertextuel, illustré dans l'exemple 2 par "*a divorce granted under this Act*".

1.2.3 Le discours judiciaire et juridictionnel

Le discours judiciaire, ou *judicial discourse*, est quant à lui centré autour de la figure de la ou du juge, qui prend sa décision pour juger une affaire, et la justifie, au terme de la procédure engagée : le juge rend une décision motivée, c'est à dire étayée de motifs. Dans cette section, nous nous concentrerons sur un type de textes

en particulier, produit dans une situation discursive précise : les motifs de jugements rendus par l'instance juridique la plus haute du Canada, la Cour Suprême du Canada (CSC), qui publie ses jugements en français et en anglais. C'est pourquoi nous parlons ici à la fois de discours judiciaire (renvoyant à la procédure judiciaire au tribunal) et de discours juridictionnel car, comme nous l'avons vu, ce sont les jugements des cours supérieures qui sont le fondement du système de *common law*, système fondé sur contrôle judiciaire (*judicial review*). Nous avons choisi la Cour Suprême du Canada car elle a une politique de diffusion de ses jugements sous forme de résumés rédigés en style clair. Nous décrivons ici le contexte et la structure générique du discours judiciaire, dans le cadre spécifique de la procédure d'appel canadienne.

Si les justiciables contestent une décision de justice rendue en première instance (c'est-à-dire lors d'un premier procès), ils ou elles peuvent faire appel de cette décision devant une cour d'appel provinciale. Une fois que cette cour d'appel a rendu son jugement, si l'une des parties concernées la conteste toujours, elle peut porter l'affaire devant la Cour Suprême. Les juges de la Cour Suprême, au nombre de neuf, statuent alors sur le cas en question, notamment sur la question juridique qui en découle. Les jugements sont ensuite rendus oralement. Généralement, un.e seul juge rend le jugement pour tous.tes les juges qui sont d'accord avec sa décision. Une autre opinion concernant l'affaire peut être rendue par un ou plusieurs juges qui ne sont pas d'accord avec la majorité (*dissenting opinion*). La décision, appelée *arrêt*, est, comme nous l'avons déjà évoquée, motivée et donc accompagnée des Motifs de jugements (en anglais *Reasons for judgments*), c'est-à-dire que les juges exposent l'argumentation qui mène à leur décision. Pour chaque affaire, l'ensemble (arrêt + motifs) est ensuite retranscrit par écrit et notamment publié en ligne sur le site de la CSC.

La structure du genre discursif du jugement est, comme celle des textes de lois, hautement contrainte (Maley 1994), et généralement composée des étapes suivantes, présentées ci-dessous et accompagnées d'exemples de la CSC pour les illustrer :

- Rappel des faits et événements de l'affaire (appelée *cause* en français canadien). Pour les jugements en cours d'appel, sont également incluses ici les décisions prises en première instance par les cours inférieures.
 - (4) Au printemps 2012, M. Wong a été accusé d'un chef de trafic de cocaïne auquel il a finalement plaidé coupable.
- Question de droit posée par l'affaire.
 - (5) At issue in this case is whether the Canadian Human Rights Tribunal ("Tribunal") erred in finding that legislative enactments cannot be challenged pursuant to 5 of the Canadian Human Rights Act.

- Motifs (*reasons* du jugement) : interprétation de la règle de droit et démonstration du raisonnement menant à la décision.
 - (6) For the reasons that follow, I would allow Mr. Ewert's appeal in part, and declare that the CSC did in fact breach the obligation in 24 of the CCRA.
- Conclusions et décision : "verdict" de la Cour
 - (7) We are therefore of the view that a new trial is not warranted. The appeal is dismissed. Judgment accordingly.

Les exemples illustrent plusieurs différences avec les textes législatifs décrits plus haut. Tout d'abord, on notera le caractère beaucoup plus personnel du Jugement, rédigé à la première personne du singulier. Les juges s'expriment à la première personne du singulier ou du pluriel³. De plus, si le genre législatif est principalement prescriptif et déclaratif, on voit que le jugement, à l'inverse, contient plusieurs modes discursifs en lui-même (Maley 1994). Partiellement narratif, en particulier dans la partie de rappel des faits, les motifs du juge présentés contiennent également des passages argumentatifs, puisque les juges expliquent et justifient leur décision, et argumentent pour expliquer les motifs de leur décision, et/ou d'éventuelles divergences s'il n'y a pas l'unanimité parmi les neuf juges. Le jugement dans le système de *common law* est également prescriptif et performatif, tout comme le sont en droit français les arrêts des cours d'appel et de la Cour de cassation. En effet, les juges rendent non seulement une décision qui prend effet au moment où ils l'énoncent, mais ils proclament également la loi, car leur décision fera précédent pour les cours inférieures pour prendre une décision dans des cas similaires. Ce genre discursif est par ailleurs également caractérisé par une forte intertextualité. Il contient en effet des références non seulement aux textes législatifs, comme on le voit ci-dessous, mais aussi des décisions précédentes prises par les cours d'appel et, en particulier, la Cour suprême elle-même.

- (8) **Section 489.1 of the Criminal Code** requires that the seizure and detention of property by the police be reported to a justice "**as soon as is practicable**".
- (9) Fundamentally, reviewing courts cannot simply "**pay lip service to the concept of reasonableness review while in fact imposing their own view**" : **Dunsmuir , at para 48**.

Le jugement de la Cour Suprême présente donc une forme d'hybridité discursive, puisque plusieurs modes discursifs le composent : il est tour à tour narratif, avec le

3. En procédure française, dans les tribunaux de première instance, le juge parle de lui-même à la première personne du pluriel.

rappel des faits, explicatif, puisqu'il expose le problème juridique posé par le cas (en droit), argumentatif, et normatif :

Le jugement est le lieu d'une rencontre (...). La nécessité de se justifier en fait et en droit y introduit plusieurs langages, celui du fait, celui de la preuve, celui du droit et celui de la logique. (...) Le discours juridictionnel est plein de discours incorporés. (Cornu 2005 : 335-366)

En termes de modalité, il est également pluriel, puisque les jugements de la Cour suprême sont souvent prononcés oralement le jour de l'audience avant d'être couchés sur le papier sous la forme de textes, appelés opinions ou motifs. Tout comme les rédacteur.ices des textes législatifs s'adressent à un lectorat multiple (les justiciables, mais aussi et surtout aux professionnels du droit), les juges de la Cour Suprême, lorsqu'ils rendent leurs jugements, ont donc également en tête un lectorat plus large que les parties concernées, puisque leur décision et leur argumentation a une incidence sur un public plus large (Maley 1994). Le site de la CSC explique ainsi :

L'autorisation est accordée par la Cour si, par exemple, l'affaire comporte une question d'importance pour le public ou une question importante de droit (ou une question mixte de droit et de fait) qui justifie l'intervention de la Cour. La décision de la Cour d'autoriser ou non un appel repose sur son appréciation de l'importance pour le public des questions de droit soulevées. (Cour Suprême du Canada 2017)

C'est donc même la pertinence de l'affaire pour l'ensemble des citoyen.es qui guide le choix des autorisations d'appel par la Cour. Ces cas, étudiés et jugés par la Cour, font ensuite jurisprudence pour l'ensemble des cours inférieures, ce qui constitue véritablement la substance du système juridique de *common law* (Maley (1994), Deumier (2019)).

1.2.4 Le discours administratif

A ces deux grands sous-registres du discours juridiques vient s'ajouter un troisième, le discours administratif, "celui que connaît le grand public" et qui est généralement assimilé à la loi, d'après (Gémar 1981). Il s'agit du discours émis par les gouvernements et administrations aux niveaux national et local, sous la forme de divers genres discursifs, comme les documents servant à communiquer avec les institutions (formulaires, etc.) mais également la notification, acte qui fait connaître un fait ou une décision juridique ou administrative :

C'est la forme ordinaire de la correspondance dans les relations d'affaire ou la pratique administrative. (Cornu 2005 : 243)

Très codifiée, la notification doit contenir le nom de la personne concernée, de la personne ou institution émettrice, sa finalité, et inclut souvent une ou des références à un ou plusieurs textes législatifs. Ce type de texte peut être par exemple un acte d'huissier.ère, ou une lettre d'une notifiant une personne, d'une décision la concernant, comme l'ouverture ou la fermeture de droits sociaux. La notification est appelée *réceptice* (Termium 2015), c'est-à-dire qu'elle est unilatérale et que sa finalité est d'être notifiée à l'intéressée.e. ; pour produire des effets juridiques, elle doit être portée à la connaissance de son ou sa destinataire. Si on ne peut s'assurer que le ou la concerné.e a eu connaissance de l'acte de justice (par exemple convocation à une audience, copie d'un jugement), la personne peut par la suite faire opposition à une décision.

D'autres genres discursifs issus de la communication entre les instances et les justiciables sont quant à eux informatifs, c'est-à-dire destinés à expliquer le droit au grand public, et disponibles sous forme de brochures ou, de plus en plus souvent, en ligne. Ils sont présentés en détail à la Section 3. C'est ce discours qui est au cœur de notre travail de recherche et de la constitution de notre corpus.

1.3 *How To Speak Legalese* : Traits linguistiques et rhétoriques du discours du droit

Bien sûr, puisque plusieurs sous-registres co-habitent sous l'appellation "discours du droit", des divergences sont palpables entre le style législatif, le style juridictionnel ou le style administratif. Cependant, se retrouvent à travers ces différents types de discours certains traits transversaux au registre juridique, des "points communs" liés à une fonction commune à tous ces registres (Cornu, 2005 : 213). Ces traits discursifs forment le style de ces discours, qu'on peut qualifier de style juridique :

Legal style, then, is related to use, function and activity. What is being done has a direct impact on word choice, and this in turn produces the distinctive register that can be measured and observed in frequency lists and analysis of extracts of legal text and talk (Coulthard et al. 2016 : 41).

Dans cette section, nous nous attachons donc à décrire les caractéristiques discursives et lexico-grammaticales qui donnent au discours juridique son style immédiatement reconnaissable. Nous nous appuyons particulièrement sur trois auteurs pour décrire les types de discours juridiques et leur saveur distinctive choisis pour cette

étude. Le premier est Maley (1994), qui propose une synthèse des trois grands types de discours de la *common law* en adoptant une approche systémique et fonctionnelle, qui relie leurs traits lexico-grammaticaux centraux d'un type de texte à sa situation discursive générale (*field*) et aux relations, rôles et objectifs des participants humains ou institutionnels (*tenor*). Le second auteur est Cornu (2005), qui fait une typologie similaire des discours juridiques francophones tout en proposant une véritable linguistique juridique, notamment une caractérisation du vocabulaire juridique. Enfin, nous nous appuyons également sur les ouvrages de Tiersma, en particulier Tiersma (2000). Ces auteurs s'attachent en effet à synthétiser caractéristiques du discours juridique selon la littérature ; ils décrivent donc différents sous-registres et genres juridiques, par exemple les contrats et testaments, mais également les textes de lois et décisions de justice. Nous nous appuyons sur leurs synthèses, mais également sur les recherches récentes (et moins récentes) en linguistique de corpus et en anglais de spécialité pour identifier les traits discursifs pertinents pour notre travail.

1.3.1 Perceptions et réputation du langage du droit

Comme nous l'avons déjà évoqué en introduction, le discours juridique souffre d'une mauvaise réputation, puisqu'il est souvent considéré par les non-spécialistes comme composé de formes "bizarres et inaccessibles" (Maley 1994 : 14), ou encore "délibérément obscur voire par moments complètement impénétrable" ("*deliberately obscure and at times totally impenetrable*") (Coulthard et al. 2016 : 39) et archaïque. Si, comme nous le verrons, ce sentiment fort d'obscurité et d'incompréhension est en partie dû à des caractéristiques linguistiques et discursives, il faut souligner les multiples relations de pouvoir qui s'exercent entre les institutions et leurs usagers dans les situations discursives mentionnées plus haut, par exemple dans l'énoncé de la loi, le rendu d'un jugement ou la notification d'une décision administrative.

Power disparities and hyper-formality are produced by the essentially controlling nature of the legal system. (Gibbons 1995)

Les discours juridiques sont en effet caractérisés par des inégalités fortes entre les participants : ceux qui exercent le pouvoir sont non seulement ceux qui maîtrisent le lexique et la phraséologie spécialisées, mais également ceux qui ont le plus droit à la parole, et qui contrôlent le sujet (Maley, 1994). Cutts (2008), évoque la dimension oppressive de la langue administrative et juridique en la reliant à la violence politique, puisqu'elle peut être utilisée pour mettre les minorités au ban de la vie

politique et judiciaire⁴. Dans les différentes situations discursives du droit évoquées plus haut, les rapports de pouvoir sont ritualisés et exprimés par des formules extrêmement figées et hautement soutenues, qui sont le produit de la construction socio-historique du système juridique. La langue spécialisée est non seulement objet et moyen du droit, elle est aussi un marqueur d'appartenance à la communauté de discours spécialisée, et fondatrice de l'identité professionnelle Turfler (2015) des acteurs et actrices du droit, des législateurs aux juges en passant par les membres des gouvernements locaux. Les membres d'un tel groupe social spécialisé, par exemple médecins ou avocat.e.s, se distinguent du "grand public" non seulement par l'acquisition de diplômes ou certificats attestant d'un certain niveau de connaissances ou de pratiques spécialisées, mais aussi par l'utilisation d'une terminologie et la capacité à comprendre et à produire des types de textes spécialisés. La difficulté serait donc liée à la multiplicité des lectorats visés. Tout comme les avocat.e.s s'adressant aux témoins, les textes de lois et jugements des cours supérieures ne s'adressent pas exclusivement, ni même principalement, nous l'avons vu, aux justiciables (Maley 1994). Pour certain.e.s justiciables, l'hétérogénéité énonciative (Preite 2018) de la langue du droit rend ce discours obscur et vise l'exclusion des non-experts, celles et ceux qui n'appartiennent pas à la communauté de discours. La dimension ritualisée et conservatrice (voire archaïsante) du discours aurait donc pour but de maintenir les relations de pouvoir existantes - c'est en tout cas ainsi qu'elle est perçue (Maley 1994). Les études psycholinguistiques portant sur la perception de textes à composante juridique, par exemple les conditions générales de vente pour des produits financiers, suggèrent que l'attitude des lecteurs.ices vis-à-vis de ces textes perçus complexes est teintée de découragement. Certain.e.s participants arrêtent par exemple la tâche demandant la lecture du texte, le lisent en diagonale, ou signalent une difficulté de compréhension et le besoin de lire le texte plusieurs fois (Rossetti et al. 2020). C'est notamment pour répondre à ces critiques que le Plain Language Movement (PLM) est né. Cependant, une approche fonctionnaliste adoptée notamment par Coulthard vient nuancer cette opposition entre initié.e.s et profanes :

[...] as Halliday (1994) points out, language is the way it is because of what it has to do. On the one hand we can argue that such language is difficult to understand and therefore distances and disadvantages the lay participant, but an alternative functional perspective is that the formulaic formality is part of the way the participants orient what is going on. (Coulthard et al. 2016 : 37)

4. " *The fact that obscurity could be oppressive had been shown in the 1960s when civil-right activists challenged a Louisiana statute requiring black (African-American) people to interpret difficult passages from the constitution before they were allowed to vote.*" (Cutts 2008 : xiii)

Ce serait donc justement pour signaler formellement le début d'une procédure judiciaire ou d'un acte juridique formel que sont faits les usages spécialisés de la langue du droit. Les traits discursifs et terminologiques perçus comme complexes par les usagers sont au contraire vus par les spécialistes comme la manière la plus claire et précise d'exprimer et de réaliser la technicité du droit, qui provient elle-même de ses fonctions spécifiques. Nous présentons maintenant en détail ces traits linguistiques et discursifs.

1.3.2 La terminologie et le lexique juridiques, une clé vers le discours du droit

Bien que la terminologie et le vocabulaire juridiques soient bien sûr construits et pleinement intégrés au discours juridique, ils font l'objet d'une description particulière car, d'après certains auteurs, la terminologie, ou encore le "jargon" juridique, est très certainement la seule caractéristique linguistique commune à tous les genres du langage du droit, qu'il soit parlé par les juges, les avocats ou les gouvernements (Tiersma, 1999 : 136 ; Cornu, 2005 : 60).

Terminologie juridique et polysémie

Dans le Tableau 1.1 sont présentés des exemples de termes juridiques, qui illustrent les particularités de cette terminologie. Ils sont tirés d'un glossaire de termes juridiques anglophones publié par le Ministère de la Justice canadien.

TABLEAU 1.1 – Exemple de termes juridiques anglophones et leurs définitions (Imprimeur de la Reine pour l'Ontario, 2016)

Terme	Définition
Justice	1. The principle of giving every person his or her due. 2. A judge.
Certiorari	A remedy used by a superior court to quash an order or decision of a lower court or administrative body made without jurisdiction.
Child	For most legal purposes, a person under 18 years of age without the legal rights and responsibilities of an adult.

On distingue plusieurs types de termes, selon qu'ils ont ou non un sens à la fois dans les contextes juridiques et dans les contextes dits généraux, non-spécialisés. Les

termes d'appartenance juridique exclusive "n'ont de sens qu'au regard du droit" et n'ont aucun "sens extrajuridique, même métaphorique" (Cornu 2005 : 62). Certains de ces termes sont peu employés ou seulement par certaines professions juridiques ou à certaines étapes de la procédure judiciaire. Dans le Tableau 1.1, le terme *certiorari* est l'un de ces termes exclusivement juridiques. Il désigne " l'ordre d'une cour supérieure enjoignant à un tribunal inférieur de lui communiquer les pièces d'une affaire dont il s'est saisi afin d'évoquer la cause avant jugement ou de réviser le jugement s'il est déjà rendu" (Définition du *Juridictionnaire*, Centre de traduction et de terminologie juridiques (CTTJ), Faculté de droit, Université de Moncton, 2015). Les termes de ce type (aux emplois exclusivement juridiques) sont souvent perçus par les non-spécialistes comme "hermétiques", "au sens caché", et ils ont un sens juridique "de précision" (Cornu 2005 : 67), c'est-à-dire qu'ils désignent le plus précisément possible un élément ou concept du droit. Il est à noter que ce terme est spécifique à un territoire donné, puisqu'il est employé dans le contexte du Canada (mais également des Etats-Unis) ; en Nouvelle-Zélande et au Royaume-Uni, ce terme a été remplacé lors d'une réforme du système judiciaire ; on voit donc ici en action la spécificité territoriale du droit, et de sa terminologie. Ce terme illustre par ailleurs une autre caractéristique de la terminologie juridique : sa tendance à utiliser des emprunts. En contexte anglophone, ces emprunts viennent historiquement du français⁵ et du latin (Richard 2018), comme c'est le cas ici. Certains termes ont été traduits du latin, tandis que d'autres sont restés sous leur forme originale, et ont été intégrés tels quels.

L'autre catégorie de termes est ceux qui ont une *double appartenance*, c'est-à-dire qu'ils ont un sens dans la langue dite générale et un sens dans la langue juridique. Certains de ces termes sont des mots-clés, "porteurs des notions fondamentales du droit", qui expriment les notions, catégories et opérations juridiques de base qui sont au cœur du système juridique : en anglais *court*, *justice*, *judge*, *trial*, en français *juge*, *tribunal*, *témoin* (toujours selon Cornu 2005 : 62). Ces mots-clés sont facilement compréhensibles par les justiciables parce qu'ils désignent ou sont utilisés par beaucoup d'utilisateurs et font partie de la vie juridique quotidienne, bien qu'ils aient également un sens spécialisé précis. Ils ont par ailleurs essaimé dans le langage dit général, puisqu'ils peuvent être utilisés dans des locutions et expressions où ils prennent un sens figuré (Cornu, 2005 : 71). Le terme "justice" illustre bien cette double appartenance (Tableau 1.1). Il renvoie, dans ses emplois généraux comme

5. Jusqu'au 18ème siècle, où plusieurs lois du Parlement britannique imposent l'utilisation de l'anglais dans la pratique juridiques au Royaume-Uni, le français était la langue du discours juridique (Richard 2018)

dans ses emplois spécialisés, au même principe fondamental de justice. Mais il est également un exemple de ce qu'on peut appeler la **polysémie interne**, c'est-à-dire qu'il a deux sens en droit. Le premier est donc, on l'a vu, le concept sur lequel se fonde le système judiciaire ; l'autre sens, dans les systèmes de *common law*, en français canadien comme en anglais, est un titre donné aux juges des cours supérieures, notamment des cours suprêmes.

Une partie des termes juridiques réalisent donc le phénomène de **polysémie externe**, c'est-à-dire qu'ils ont un sens (ou plutôt un usage) spécialisé en contexte juridique et un autre sens (ou usage) dans la langue générale. Le terme, en contexte juridique, devient chargé d'une "spécificité juridique" et "prend un sens très spécial et technique" (Cornu 2005 : 77). Dans le Tableau 1.1, un autre exemple est le mot *child*. Si, dans ses emplois généraux, ce mot renvoie de manière assez vague à toute personne avant l'adolescence, il a en droit une définition beaucoup plus précise (renvoyant aux personnes mineures). Ce sont notamment ces termes, en plus des emprunts au latin, qui peuvent prêter à confusion lorsque le justiciable les rencontre. L'exemple du terme "succomber" est encore plus parlant, puisqu'on voit ici une vraie rupture de sens entre l'usage général et l'usage juridique (Cornu 2005 : 78).

- (10) Le demandeur en cassation qui **succombe** dans son pourvoi ou dont le pourvoi n'est pas admis peut, en cas de recours jugé abusif, être condamné à une amende civile (...). (FR-Lois)

Pour un profane n'ayant aucune connaissance juridique, le mot *succomber* peut laisser croire que le ou la demandeur meurt pendant son appel ; dans ce contexte, il signifie en fait *perdre son procès*. Les termes du droit sont donc caractérisés par une forme de polysémie, et par des emplois et sens spécifiques. À part pour les termes d'appartenance exclusive (qui n'ont de sens qu'en droit) on peut conclure avec Richard que la terminologie juridique est donc davantage une question d'*usage* juridique que de *sens* juridique (Richard 2018).

Lexique juridique : registre soutenu, abstraction et imprécision

D'autres traits lexicaux et non terminologiques sont caractéristiques du lexique juridique. Tout d'abord, comme évoqué plus haut, le discours du droit est particulièrement soutenu ; il contient donc des éléments lexicaux correspondant à "un parler distingué, élevé, parfois désuet" (Collette et al. 2002), parfois qualifiés de littéraire

(Cornu 2005 : 246). En anglais, un exemple souvent cité est celui des adverbes *hereinafter* et *thereto*. Ces adverbes, considérés comme archaïques et extrêmement soutenus, sont des adverbes complexes, c'est-à-dire formés de plusieurs prépositions ou adverbes, auxquels sont préférés, dans les textes juridiques plus récents, leur équivalent courants, respectivement *from this point onwards* (*within this text*) et *to that* (Commission 2015). Ces adverbes sont d'ailleurs des adverbes complexes, c'est-à-dire des adverbes composites, ou locutions adverbiales. En anglais comme en français, ce registre élevé est également réalisé à travers l'emploi d'items lexicaux marqués comme soutenus dans la langue dite générale : Galonnier (1997) évoque par exemple le choix de la préposition *save* à la place *except* ; on peut citer également des adjectifs comme *deceased* plutôt que *dead*⁶, ou encore *novel* plutôt que *new*⁷.

Une autre caractéristique importante du vocabulaire juridique est l'utilisation de mots et expressions génériques, par exemple *person* ou en français *chacun* pour désigner toute personne susceptible d'enfreindre une loi. C'est la visée normative des textes renvoyant à la loi et à la justice qui explique l'emploi de mots et expressions génériques (Richard (2018), Tiersma (2000) : 79), qui renforcent leur caractère abstrait. Ces items génériques sont également accompagnés d'éléments lexicaux apparaissant comme subjectifs ou vagues, qui permettent aux rédacteur.ices d'une loi de l'adapter à différents contextes et juridictions, tout en restant assez flexibles pour anticiper des évolutions éventuelles des normes sociales (Tiersma (2000) : 80). C'est le cas par exemple de *reasonable* ou de *justes* dans les extraits cités ci-dessous :

- (11) A delegate or a constable executing a warrant may use such **reasonable** force as may be necessary to do so. (EN-LEX)
- (12) Toutefois, le tribunal peut, s'il y a de **justes** motifs, prononcer l'adoption lorsque la différence d'âge est inférieure à celles que prévoit l'alinéa précédent. (FR-LEX)

Ces éléments lexicaux vagues et apparemment subjectifs, sont appelés *weasel words* par Richard (2018), car ils sont d'après l'auteure délibérément ambigus. Ils correspondent d'après elle au degré d'imprécision nécessaire à la formulation de la loi pour qu'elle puisse s'appliquer à des situations et cas divers, et pour qu'elle puisse être interprétée à la discrétion des magistrats. Ces mots et expressions imprécis en

6. Par exemple, (...) *the personal representative of the **deceased** person or other successor in title to property of the **deceased** person is entitled to assert the privilege against third parties* (Texte législatif néo-zélandais).

7. Par exemple *The law would stagnate in the absence of creative and **novel** legal argumentation.* (Jugement de la CSC)

viennent cependant à poser des problèmes d'interprétation pour les juges, qui sont souvent au cœur de l'argumentation dans les jugements rendus. Les magistrat.e.s doivent décider selon une contextualisation précise si un mot présent dans le texte de loi correspond à la situation dans une affaire jugée (Richard 2018), comme on le voit dans un extrait de jugement de la Cour Suprême du Canada ci-dessous :

- (13) By contrast, in a judicial review of a grievance proceeding, the usual question is whether the CSC's decision based (...) on information in the inmate's file (including test results) was **reasonable** (extrait d'un jugement de la CSC).

Ici les juges doivent donc statuer sur le caractère raisonnable d'une décision antérieure, en argumentant selon leur propre interprétation de ce qui est raisonnable dans le texte de loi. La question de droit posée par une affaire est donc intrinsèquement liée à la formulation de la loi, et confirme que langue et droit sont indéniablement liés. L'interprétation des textes, et en particulier de l'intention des législateur.ices et dans la jurisprudence est un enjeu crucial pour les juges, comme le montre l'exemple 14, où le magistrat s'arrête longuement sur les définitions de l'adjectif *particular* et cite un dictionnaire de langue générale.

- (14) By applying Philip Morris's interpretation of "**particular**" in 2 to the same text in 2, the latter would read : "it is not necessary to identify identifiable individual insured persons". It seems unlikely that the Legislature intended to say this. The ordinary meaning of the word "particular" is distinct or specific. This is consistent with the **Oxford English Dictionary** (online) which defines "particular" as meaning "one among a number single; distinct, individual, specific" (emphasis added). This definition supports the view that the databases – even once anonymized – fall within 2's scope as comprising the "health care records and documents of", and the "documents relating to the provision of health care benefits for", each distinct and specific individual in British Columbia. (Jugement de la CS du Canada)

Ces éléments lexicaux vagues remplissent ainsi un besoin de la loi, qui doit parfois rester vague pour être assez générique, mais ils entraînent aussi des problèmes d'interprétation. Dans les Jugements de la CSC, rendus oralement et rédigés personnellement par un juge en son nom ou au nom de plusieurs juge du même avis, le lexique compte également ces termes généraux et vagues ; mais, à l'inverse des textes législatifs, on y trouve également des mots subjectifs bien plus explicites que les mots vagues que nous venons de citer (qui sont subjectifs parce qu'ils doivent être interprétés par un.e juge). On peut en y effet trouver des méta-commentaires

sur l'affaire qui est jugée, montrant un positionnement émotionnel et personnel du ou de la juge vis-à-vis de l'affaire⁸ :

- (15) This case is undeniably **tragic**. Indeed, **few tragedies are worse** than the death of a child.

Contrairement aux genres législatifs, les jugements peuvent donc contenir des marqueurs d'attitude concernant les faits. Des études ont montré que les opinions des juridictions supérieures comprennent également des marqueurs d'évaluation des arguments et décisions des cours inférieures, par exemple dans les opinions de la Cour Suprême des Etats-Unis (Goźdź Roszkowski et Pontrandolfo 2013).

1.3.3 Caractéristiques discursives du discours du droit

Outre par la terminologie et le lexique, les discours du droit sont caractérisés par différents traits lexico-grammaticaux et discursifs qui ont été abondamment décrits. Nous présentons d'abord des extraits de textes juridiques qui servent à illustrer ces caractéristiques discursives du registre juridique, en anglais dans le Tableau 1.2 et en français dans le Tableau 1.3. Bien qu'ils n'appartiennent pas aux mêmes genre discursif (législatif vs. judiciaire), ni au mêmes contextes culturels ou à la même langue, on trouve dans chacun de ces extraits des traits transversaux présentés dans la littérature comme typiques du discours juridique anglophone comme francophone, au-delà des frontières (Galonnier 1997). Comme nous le verrons dans la description de ces traits grammaticaux et discursifs, ils fonctionnent souvent en miroir des recommandations pour un style clair présentées au Chapitre 2.

TABLEAU 1.2 – Extraits du discours juridique anglophone

Discours législatif	Discours judiciaire
(a) A person A discriminates against a woman <u>if</u> , in the period of 26 weeks beginning with the day on <u>which</u> <i>she gives birth</i> , A treats her unfavourably <u>because</u> she has <i>given birth</i> . (texte de loi)	(b) The next question is <u>whether</u> the LSBC's decision not to approve TWU's law school limits the ability of TWU's community members to act in accordance with these beliefs and practices in a manner that is more than trivial or insubstantial (Amselem, at para 74 ; Ktunaxa, at para 68).(Jugement de la CSC)

8. Ce qui n'est pas du tout le cas en procédure française, où le juge peut rendre compte de l'attitude ou des sentiments des parties mais ne délivre pas d'appréciation personnelle.

TABLEAU 1.3 – Extraits du discours juridique francophone

Discours législatif (France)	Discours judiciaire (Canada)
(c) Constitue, au sens des articles 311-4 (...) et 311-10, un vol suivi de violences le vol à la suite duquel des violences ont été commises pour favoriser la fuite assurer l'impunité d'un auteur ou d'un complice. (article de loi)	(d) Chacun a droit à la vie, à la liberté et à la sécurité de sa personne ; il ne peut être porté atteinte à ce droit qu'en conformité avec les principes de justice fondamentale. (décision de la CSC)

Redondance

Les exemples ci-dessus montrent une tendance à la répétition, puisqu'un ou plusieurs groupes nominaux sont répétés au moins une fois dans ces courts extraits et non pas remplacés par des pronoms : *person* (a), *TWU's* (b), *vol*, *violences* (c). Les mots dans ces exemples sont répétés, afin d'éviter toute ambiguïté, ce qui peut contribuer à allonger considérablement la longueur des phrases (Tiersma (2000) : 59-60). Cette tendance à la répétition des mots et des collocations et unités phraséologiques peut conférer un style d'une grande redondance aux textes juridiques.

Complexité syntaxique et longueur des phrases

La complexité syntaxique du discours juridique, et du discours législatif en particulier, est particulièrement célèbre et immédiatement évidente, peut-être même davantage que la terminologie, selon certains auteurs comme Maley (1994). Cette complexité syntaxique se réalise par :

- Un haut degré de subordination, appelée dans certains cadres d'analyse, comme celui de la Grammaire Systémique Fonctionnelle, *hypotaxe*. Dans les textes juridiques, on observe en particulier la présence au sein d'une même phrases de propositions conditionnelles en *if* (a) ou *whether* (b) ainsi que de des propositions relatives et autres propositions circonstancielles (*adjuncts*) venant préciser et qualifier les conditions d'application de la loi, ou la question de posée par la situation en question. Ce sont les structures des genres discursifs qui expliquent en partie cette caractéristique syntaxique, comme nous l'avons vu plus haut pour le discours législatif. (*where...shall / lorsque/ si...*).
- Des phrases longues : Bhatia (1994) montre que les phrases d'une section ou d'un article de loi anglophone font en moyenne 200 mots, contre 20 pour la longueur moyenne des phrases dans les corpus de langue générale. Cette longueur est due à la fois au haut degré d'hypotaxe évoqué ci-dessus, mais

également à la coordination de plusieurs propositions, appelée *parataxe*. Les définitions des termes employés, et les énumérations d'éléments lexicaux pour couvrir tous les cas d'application de la loi expliquent aussi cette tendance.

- L'utilisation de prépositions complexes, c'est-à-dire de la forme < Préposition + GN + Préposition >, en particulier *for the purpose of*, ou dans les Tableaux 2.2 et 2.3 "*à la suite duquel*", "*en conformité avec*", "*in accordance with*". Ces prépositions complexes, d'après les juristes, ne sont pas synonymes de leurs versions simples : elles réduisent le risque d'ambiguïté (Bhatia 1994) ; elles peuvent à l'inverse être source de confusion et ajouter à la complexité syntaxique pour les non spécialistes.
- L'emploi d'expressions binomiales et multinomiales, c'est-à-dire une séquence contenant mots d'une même catégorie grammaticale (souvent des GN) coordonnés, par exemple *à la vie, à la liberté et à la sécurité* dans l'exemple (d) du Tableau 1.3 (Gustafsson 1984).

Certaines structures syntaxiques sont spécifiques à des langues et à des contextes. Les articles de loi de France, comme l'exemple (c), présentent souvent une antéposition du groupe verbal, donc une inversion de l'ordre typique Sujet-Verbe dans "*Constitue un vol...*". Cet ordre non canonique des mots forme un emploi spécifique de la langue dans un contexte communicatif spécifique, ce qui est caractéristique des langues de spécialité. Vis-à-vis de la longueur des phrases et de la complexité, la littérature suggère également certaines différences dans les styles juridiques francophones et anglophones. Taylor (2003), en comparant l'adaptation d'une directive de l'Union Européenne sur les produits défectueux dans les droits nationaux anglais et français, remarque que le texte anglais ajoute des détails par rapport à la directive, ce qui se traduit par des sections beaucoup plus longues et complexes syntaxiquement. Les rédacteur.ices anglais.es précisent en effet les notions juridiques (telle que *safety* dans ce cas précis), ce que ne font ni le texte français ni la directive européenne elle-même. D'après l'auteur, cette différence provient du besoin ressenti par les premiers de réduire les ambiguïtés, pour guider les juges dans leur interprétation. Par comparaison, le style français apparaît comme plus clair et plus concis (ce qui peut paraître ironique à toute personne française non juriste ayant dû un jour consulter un texte de loi).

Tournures impersonnelles et abstraction

Le discours juridique est caractérisé par l'emploi de tournures impersonnelles, caractéristique due à la visée générique et universelle de la loi, qui doit s'appliquer

à tous.tes les citoyen.nes. Les textes juridiques, nous l'avons vu, s'adressent souvent par ailleurs à plusieurs lectorats - avocat.es et magistrat.es, justiciables et fonctionnaires - ce qui expliquerait pourquoi les tournures à la troisième personne sont préférées aux formes de première et deuxième personne ((Tiersma 2000 : 67). En anglais, dans l'extrait de texte législatif, le caractère générique est poussé à l'extrême : pour que la loi s'applique à tous.tes sans distinction, les justiciables sont désignés par une étiquette de type $\langle Person + Numérotation \rangle$, ici *a person A* et *a person B* dans le cas de la définition du concept de discrimination. La loi réfère donc à toutes personnes indistinctement et peut s'appliquer à toute une série de contextes et de situations. Cette stratégie n'est pas sans rappeler la formulation et le degré d'abstraction d'un énoncé mathématique⁹. En français, le discours législatif réalise la généralité et l'abstraction de la règle de droit par l'utilisation de la troisième personne, en particulier de formes impersonnelles en "Il" (Cornu 2005 : 276).

Dans les jugements rendus par les cours supérieures comme le Canada, en (b), on l'a vu, chaque juge s'exprime pour soi-même ou pour la majorité. L'opinion, genre à la fois explicatif et argumentatif, fait donc usage notamment des pronoms personnels de première personne (*with which I take no issue* dans le Tableau 1.2). puisque le ou la juge prend la parole pour expliquer le raisonnement ayant mené à la décision. On y trouve cependant également des tournures impersonnelles, lorsqu'il est fait référence à la loi et à des principes de justice, par exemple l'utilisation du il impersonnel ("il ne peut être porté atteinte") et du pronom générique "chacun" (Tableau 2.3, ex/(d)) pour désigner des règles de droit ou les justiciables. Le caractère normatif de la loi et de la justice impliquent un tel degré de généricité et d'abstraction, ce qui augmente la complexité, le degré de personnalisation et d'abstraction d'un texte ayant un impact sur sa lisibilité, comme nous le verrons au Chapitre 3.

Le passif dans le discours du droit

Dans le Tableau 1.3 qui présente des exemples francophones, on trouve des structures passives comme *il ne peut être porté atteinte* (d) et *des violences ont été commises* (c). Or, le sur-emploi du passif est l'une des caractéristiques souvent évoquées lorsqu'on décrit le discours juridique francophone (Cornu 2005 : 278), et également reprochées au discours juridique anglophone ((Tiersma 2000 : 75). Dans le cadre du discours juridique, on considère souvent que le passif est utilisé dans le but d'effacer

9. La similarité est ici frappante puisqu'on pourrait même reformuler cet extrait en utilisant la phraséologie des mathématiques : "Let A and B be persons. If A treats B unfavourably because he/she has given birth, then A discriminates against B."

l'Agent¹⁰ responsable d'une l'action, comme l'indique Tiersma lorsqu'il mentionne que les avocat.es et personnalités politiques y sont particulièrement attachés (il évoque notamment la désormais classique formule *Mistakes were made.* :

No doubt the possibility of leaving out the actor explains most of the profession's affection for the passive construction. (Tiersma 2000 : 76)

L'utilisation du passif, et en particulier du passif sans complément d'agent, aurait par ailleurs pour fonction de conférer le plus d'autorité et d'objectivité aux textes de lois et jugements :

Legislators and judges want their command to appear maximally objective, to give them the greatest possible rhetorical force. (Tiersma (2000) : 76)

Tiersma donne pour exemples des textes législatifs (*statutes*) produits par le Congrès des Etats-Unis et les décisions de la Cour Suprême des Etats-Unis, qui ne contiennent, dans une série de décisions évoquées par la Cour elle-même, pratiquement que des constructions passives (Tiersma 2000 : 77).

Un tel usage de la voix passive s'observe également dans le discours scientifique, et l'utilisation de ces constructions dans les deux registres a été étudiée conjointement par Seoane et al. (2006). Dans cette étude, les auteurs comparent l'évolution de certaines structures, en particulier du passif, dans un corpus d'articles de recherche (représentatifs de l'*Academic Scientific Discourse*, ASD) et dans un corpus textes législatifs britanniques, étatsuniens et australiens (représentatifs du *Prescriptive Legal Discourse*, PLD) tous deux en diachronie (années 70 vs. début des années 2000). Les résultats révèlent une baisse drastique du nombre de structure *be-passives* dans le corpus d'articles de recherche sur la période, tandis que les textes législatifs montrent une évolution plus lente, bien que plus évidente en Afrique du Sud et en Australie. Seoane et al. (2006) expliquent cette différence par la culture juridique, conservatrice et attachée à la tradition (en particulier au Royaume-Uni et aux Etats-Unis), alors que la communauté scientifique doit adapter et faire évoluer son style afin d'apparaître à la pointe de la recherche et refléter la démocratisation de la science. Le statut de pouvoir et d'autorité recherché par les rédacteur.ices de textes juridiques, d'après les auteurs, explique que l'évolution de l'emploi du passif soit plus lente dans le discours législatif que dans le discours scientifique. Ils remarquent cependant en particulier dans le sous-corpus australien une tendance à la baisse, contrairement au corpus du Royaume-Uni et des Etats-Unis. Selon les auteurs, le poids de la tradition semble donc plus fort dans ces derniers, et elle a donc une plus grande influence dans la rédaction des textes législatifs. Une étude ultérieure de Williams (2013) vient par

10. Nous utilisons ici des majuscules pour désigner le rôle sémantique d'Agent.

la suite nuancer ces résultats, puisqu'il conclut à une baisse de la fréquence du passif dans un corpus de textes législatifs issus d'Australie, mais également du Royaume-Uni. Le nombre de constructions passives passe de plus de 50% en 1980 à 26% en 2010. Cette évolution, ajoutée à d'autres changements dans le groupe verbal, comme la disparition de l'emploi de *shall*, reflète la modernisation du discours législatif grâce à l'influence du PLM. Nous revenons sur la question du passif, point central du *plain language*, dans les Chapitres suivants.

Prédicats complexes et style nominal

L'un des traits fréquemment cités comme caractéristique du discours juridique écrit est l'emploi de groupes nominaux plutôt que verbaux, ce qui résulte en un style "hautement nominal" (*so highly nominal* Crystal et Davy (1969)). En particulier, plusieurs auteurs, identifient les nominalisations comme traits du discours juridique écrit (Tiersma 2000; Diani 2019). Une nominalisation sont définies par Williams (2004) comme le phénomène où un verbe est transformé en nom (*where verbs are transformed into nouns*). En réalité, la nominalisation ne concerne pas uniquement le passage de la catégorie verbale à la catégorie nominale, puisqu'on trouve en français comme en anglais des nominalisations qui concernent d'autres catégories grammaticales (dont l'exemple prototypique est la nominalisation de l'adjectif *rich*/"riche" en *the rich*/"les riches"). Il est notable cependant que les articles et ouvrages sur le discours du droit emploient et même définissent le terme de "nominalisation" comme la transformation du verbe en nom. Ces auteurs visent en fait les structures du type *make him a payment*, par opposition au verbe simple *pay* dans la phrase *The Secretary of State may make him a payment of such amount as the Secretary of State may determine*. (extrait d'un texte de loi). Dans cet exemple (emprunté à Diani (2019) et son étude sur la langue des contrats), un verbe générique (*make*) constitue le prédicat mais c'est son complément, sous la forme d'un nom prédicatif (*payment*), qui réalise la fonction sémantique de Procès.

De nombreux termes ont été employés dans la littérature pour désigner ce type de structures, reflétant diverses analyses selon des cadres théoriques différents : "constructions à verbe support" (Giry-Schneider 1987), en anglais *light verb constructions* - le verbe étant alors considéré comme sémantiquement "faible" (*weak verb*) - ou encore "collocations verbo-nominales" (Gledhill 2008 : 47). Dans la suite de ce travail, nous emploierons le terme de **prédicat complexe** pour désigner ce type de constructions. Expriment sous forme nominale des Procès typiquement exprimés par des verbes, elles sont associées à la notion de métaphore grammaticale,

c'est-à-dire au transfert d'une catégorie grammaticale à une autre.

Dans les ouvrages portant sur le discours du droit, ces structures sont généralement considérées comme un autre moyen d'atteindre un style générique compatible avec le degré d'abstraction et de formalité de la loi et de la justice, tout en effaçant (à l'instar du passif) l'Agent du procès comme l'indique (Tiersma 2000 : 77). Ce type de prédicats complexes est également fréquent dans d'autres discours spécialisés, comme le discours médical et le discours scientifique, et sont considérés comme participant de la complexité syntaxique, comme nous le verrons au Chapitre 3.

Intertextualité

Enfin, l'un des traits discursifs qui imprègne le plus le discours juridique est l'intertextualité, à la fois explicite et implicite :

[In judicial discourse] as in legislation and courtroom discourse, all discourse is carried on against a framework not only of shared knowledge of the institutional culture, but of a knowledge a more specific kind, that is, the relevance of other institutional discourses to the particular issue" (Maley, 1994 : 47).

Les différents genres du droit sont émaillés de références textuelles sous la forme de renvois à d'autres textes législatifs ou à des précédents juridiques explicites. Les décisions de justice, en particulier dans le système de *common law*, font à la fois référence à la loi et aux décisions précédentes, comme on le voit dans l'exemple (b) du Tableau 1.2 (*Amselem, at para 74*, p. 35). Dans les textes législatifs, il est fait référence à des articles à l'intérieur du même texte de loi (par exemple *au sens des articles 311-4 et 311-10*, (c), Tableau 1.3, p. 36), mais également à d'autres textes de l'appareil législatif. Bien souvent, les nouvelles lois et décrets qui sont passées par les parlements viennent simplement modifier les textes précédents. L'appareil législatif et jurisprudentiel, caractérisé comme on l'a vu par un fort degré d'abstraction, est également instable et voué à augmenter toujours pour modifier et abroger les textes précédents. Cette inflation textuelle peut être source de difficulté non seulement pour les justiciables, mais également pour les juristes. Plus implicitement, les genres du droit doivent être lus en regard les uns des autres, comme une myriade de textes connectés entre eux par la législation et la jurisprudence (Office of the Scottish Parliamentary Counsel 2006). La difficulté provient ici, pour les non-spécialistes, des connaissances préalables nécessaires pour dérouler les références textuelles dans les textes, et, plus généralement, comprendre le système législatif dans lequel un texte s'insère.

1.3.4 La phraséologie du droit : collocations juridiques et formulaïcité

La littérature récente pointe une explosion du nombre d'études de corpus du droit qui portent sur la phraséologie (Williams 2020). En tant que sous-discipline de la linguistique, la phraséologie s'intéresse aux séquences de mots conventionnelles et aux relations lexicales, sémantiques et grammaticales entre les mots. Si en ses débuts elle portait surtout sur les expressions idiomatiques, elle a également pour objet d'études les collocations.

Ainsi, nombre d'études linguistiques sur le droit s'intéressent aux collocations, c'est-à-dire aux termes et éléments lexicaux non pas en isolation mais en lien avec leur "environnement distributionnel" (Gautier 2003). Le terme de collocation décrit une "association lexicale conventionnelle", résultat "d'une contrainte sémantique de sélection" et "qui a pour effet de restreindre la compatibilité des mots" avec un item lexical donné (Neveu 2004). Les exemples suivants illustrent des collocations liées aux termes *appel* en français et *appeal* en anglais :

- (16) Il **interjette donc appel** de sa déclaration de culpabilité au motif qu'il y a eu erreur judiciaire au sens du sousal. 686(1)(iii) du Code criminel. (Jugement de la CSC)
- (17) Ms. Magoon and Mr. Jordan each appealed their second degree murder convictions. (Jugement de la CSC)

Le mot *appel* en français entre en collocation avec *interjeter*, mais une autre collocation possible est *faire appel*, ce qui montre que deux verbes sont compatibles avec *appel*, les assemblages lexicaux restant libres dans les collocations (Neveu 2004). En anglais, dans notre exemple, *appeal* est un verbe et entre en collocation avec *conviction* ici, mais il peut également être associé à *decision*. Ces unités lexicales sont donc plus ou moins systématiquement associés à ces verbes, et c'est cette co-occurrence conventionnelle que décrit le concept de collocation (Neveu 2004). Divers genres discursifs juridiques ont été étudiés sous l'angle des collocations, par exemple les collocations contenant des expressions binomiales dans les contrats (Basanež 2017). Des études diachroniques se sont intéressées aux collocations et de la prosodie sémantique de certains mots au sein du registre juridique ; Lehto et al. (2019) étudie par exemple les collocats de *person* dans un corpus du XIXème et XXème siècle, montrant l'évolution du statut des citoyen.nes et de la conception de l'autorité dans le discours législatif. Ces études, bilingues ou monolingues, soulignent la nature hautement formulaïque du discours juridique (Breeze 2017).

En effet, la phraséologie s'intéresse également aux séquences formulaïques, qui correspondent à des séquences récurrentes d'unités lexico-grammaticales suivant un degré plus ou moins fort de figement. Les phraséologues utilisent différents termes pour désigner ces séquences récurrentes, que nous définissons ici brièvement. Plusieurs modèles se sont en effet attachés à décrire l'inextricable association entre les items lexicaux et des schémas grammaticaux et syntaxiques. Pour *appel* et *appeal*, on voit par exemple que les collocations reposent sur une construction grammaticale spécifique, puisque, en français par exemple, le nom n'est pas précédé par un article, mais il est suivi par la préposition *de* (*interjeter appel de*), tandis qu'en anglais, le verbe *appeal* prend un complément d'objet direct. En linguistique de corpus, le terme *n-gram* désigne des séquences contiguës de mots ou lemmes présents dans certains genres (Kraif et Tutin 2017), qui n'ont parfois pas de statut linguistique (il sont seulement identifiés comme des séquences de mots et de ponctuation), mais correspondent parfois à des unités phraséologiques dont on peut identifier une fonction discursive précise. Les modèles cognitivistes de la grammaire de construction (*Construction Grammar*) désignent par *construction* un mot ou groupe de mots constituant une unité grammaticale qui associe une forme à une fonction (Goldberg 2006); ces constructions n'ont à l'origine pas été théorisées pour des contextes spécialisés, cependant l'influence de la grammaire de construction est palpable et certaines unités phraséologiques présentes dans des genres spécialisés sont analysées comme des constructions (Groom 2019). La notion de *lexical patterns* et les travaux réalisés dans le cadre de la (*pattern grammar*) soulignent l'interdépendance entre les items lexicaux et les schémas grammaticaux dans lesquels ils se retrouvent (Hunston et Francis 2000 : 3), c'est-à-dire les séquences plus ou moins fixes de morphèmes et de lexèmes (Hunston et Francis 2000 : 7). En analyse de discours francophone, c'est la notion de routines sémantico-rhétoriques qui est utilisée pour analyser les séquences récurrentes, composées d'un patron syntaxique mais variable quant au plan lexical, et spécifiques à un genre discursif dans lequel elles remplissent une fonction rhétorique et discursive propre, (Kraif et Tutin 2017). Un exemple d'une telle routine serait *au sens des articles X et X* qui apparaît dans les textes législatifs francophones, et remplit une fonction métatextuelle "renvoyer à une autre partie du texte". Dans la suite de ce travail, nous utilisons le concept de schéma lexico-grammatical, qui se rapproche à la fois des *lexical patterns* et des routines sémantico-discursives. Issu du modèle de la GSF, un schéma lexico-grammatical (schéma LG) est caractérisé par plusieurs propriétés : il s'agit d'une séquence récurrente d'items lexicaux et grammaticaux, prévisible et productive et remplissant une fonction pragmatique et rhétorique dans un registre ou genre. Un schéma LG est composé d'un pivot, c'est-

à-dire un ou plusieurs éléments obligatoires, et d'un ou plusieurs éléments variables. La variabilité et la discontinuité éventuelle de ces éléments permettent à ces schémas d'être imbriqués dans d'autres schémas, créant des "chaînes discursives" (Gledhill et al. 2017) qui sont la pierre angulaire des discours spécialisés (Gledhill et Kübler 2016). En Nouvelle-Zélande, les textes de loi débutent par une phrase qui peut être considérée comme un schéma LG : *This Act may be cited as the [Nom de la loi] Act*. Le pivot est discontinu et il est particulièrement long et est formé par une proposition passive, tandis que l'élément variable est le nom de l'Act. La fonction de ce schéma est ici une fonction métatextuelle liée à l'intertextualité du discours juridique : il sert à indiquer comment le texte est intitulé, en particulier pour qu'il puisse être cité par les magistrat.es et dans d'autres textes législatifs ou administratifs. Or, le discours du droit, de par sa nature conservatrice et standardisée, est particulièrement propice à des études de schémas phraséologiques. De nombreuses formules rituelles servent à ouvrir ou fermer certaines situations de communication en contextes juridiques, oraux ou écrits. Une audience en France et à la CSC commence par exemple toujours par la formule "*La Cour*", qui précède l'arrivée des juges et indique aux personnes présentes qu'elles doivent se lever. D'autres genres écrits du discours du droit, comme les testaments et les contrats, contiennent des phrases (mais également des paragraphes entiers) qui sont pas (ou peu) modifiées à chaque édition d'un nouveau texte, ce qu'on appelle des *boilerplates*, ce dont le schéma LG cité ci-dessus est un parfait exemple :

Much of this is what those in the legal profession call **boilerplate** : a standard provision that is routinely¹¹ added to a particular type of document. (Tiersma 2000 : 59)

S'appuyant sur ces différentes approches de la phraséologie, nombre d'études du discours juridique se sont donc penchées sur ces séquences phraséologique récurrentes afin d'éclairer les fonctions discursives de la phraséologie du droit (Goźdz-Roszkowski et Pontrandolfo 2017), la formulaïcité étant considérée comme l'un des traits les plus typiques et les plus frappants du style juridique (You 2019). Elle a par ailleurs été étudiée en contexte multilingue, notamment dans les arrêts de la Cour de justice de l'Union Européenne (Trklja 2017), You (2019), ou dans le contexte de la transposition de directives européennes au droit national, par exemple polonais (Biel 2017). Pour ce qui est du discours oral, Szczyrbak (2017) montre le rôle crucial des motifs phraséologiques dans la construction discursive de la notion de preuve pendant les interrogations au tribunal. Mazzi (2017) identifie, à partir des ngrams, les fonctions

11. Cette citation fait donc bien écho aux notions de routines.

rhétoriques de plusieurs schémas phraséologiques dans un corpus de décisions de la Cour Suprême d'Irlande concernant des affaires liées au droit européen. Observer ce qui reste de la phraséologie juridique après simplification est l'un des points d'attention de ce travail, et nous revenons en particulier sur le concept de schéma LG dans le Chapitre 8.

Dans cette section, nous avons présenté une synthèse des caractéristiques principales du discours du droit, transversales aux langues et aux type de discours. Ces différentes caractéristiques, faites de complexité syntaxique, de redondance et d'abstraction, en interaction avec des spécificités lexicales et terminologiques, en font l'un des discours spécialisés les plus reconnaissables mais également perçu comme étranger. C'est pour cela que ce registre donne lieu à un discours de diffusion pour les non-spécialistes, que nous présentons maintenant.

1.4 Diffuser le discours du droit : comment ouvrir le domaine spécialisé aux justiciables ?

Comme nous l'avons vu, le domaine du droit est intrinsèquement complexe pour les justiciables non spécialistes, ce qui donne lieu à des mouvements de simplification par différents acteurs juridiques et administratifs.

Avant de commencer cette partie, un point terminologique s'impose. Nous avons jusqu'ici parlé de "diffusion du droit" comme expression générique pour désigner la communication qui vise à rendre accessible aux justiciables les discours et connaissances juridiques. Nous avons fait ce choix afin de traduire l'anglais *dissemination* et d'éviter le terme problématique de "vulgarisation"¹². Le terme de médiation est parfois préféré car "vulgarisation" établit un rapport hiérarchique entre les personnes qui ont la connaissance (scientifique ou autre), et celles qui ne savent pas. On parle donc parfois de médiation, ou *knowledge mediation*, notamment de médiation scientifique. Nous utiliserons de manière interchangeable *diffusion* et *médiation juridique*. Il est à noter que le terme de vulgarisation est cependant employé par les institutions elles-mêmes.

12. Pour une discussion détaillée de cette question, voir (Cassen 2010 : 183) : "S'il est bien commode, le terme de vulgarisation, et, plus encore, la réalité qu'il recouvre, posent autant de problèmes qu'ils paraissent en résoudre : ils renvoient à la relation entre les producteurs de connaissance scientifique et leurs concitoyens (...)."

1.4.1 Des acteurs divers

La diffusion du droit est prise en charge par plusieurs types d'acteurs, ayant différents degrés de légitimité concernant la prescription de normes et de recommandations linguistiques. Raus, dans son livre sur la terminologie portant sur l'égalité entre les genres en contexte multilingue, distingue deux types d'acteurs : les acteurs institutionnels et les associations et réseaux (Raus 2014).

Dans le cadre de la diffusion du droit, les acteurs institutionnels sont donc les institutions du domaine administratives et judiciaires. A l'échelle nationale et locale, les sites gouvernementaux publient notamment un grand nombre de documents expliquant leurs droits et démarches aux usagers, afin de guider les justiciables dans les méandres des procédures légales et administratives décidées au niveau des tribunaux, mairies ou administrations locales (*counties*). Les acteurs institutionnels des pays anglophones qui nous intéressent sont notamment le Ministère de la Justice de Nouvelle-Zélande, le gouvernement britannique, la Cour Suprême du Canada. Ces acteurs ont été choisis car ils affichent le style clair comme un principe de rédaction pour s'adresser au lectorat non-spécialiste dans leur contenu en ligne :

Plain English is mandatory for all of GOV.UK. (Government Digital Service 2016)
 Make your content easy to understand with plain language. (Digital Government New Zealand 2020)
 Cases in Brief are short summaries of the Court's written decisions drafted in reader-friendly language, so that anyone interested can learn about the decisions that affect their lives. (Cour Suprême du Canada 2022)

Le style clair est présenté comme le standard de communication entre les gouvernements et les justiciables. Il s'agit donc d'un principe politique, posé pour le discours administratif, qui correspond à une politique de transparence. La transparence étant un "ingrédient essentiel" pour maintenir la confiance du public dans les institutions politiques et juridiques, d'après les membres des institutions (Anthony 2018).

Il faut cependant souligner que la mise en place du style clair ne s'est pas faite sans résistance du milieu juridique¹³, la langue de spécialité étant, comme nous l'avons vu, un marqueur social. Les institutions qui encouragent l'utilisation du

13. Cutts (2013) rapporte une anecdote qui concerne des membres d'un conseil municipal au Royaume-Uni ; l'un d'eux s'élevait contre la simplification d'un rapport d'un conseil municipal : "This is just the way council reports are written, and if you want to be a councillor then you need to understand that. If you go to France, they speak French. Here in the council, we speak like this."

style clair doivent donc se défendre contre des accusations de simplification abusive de la langue ou de condescendance :

Plain language is not unprofessional writing or a method of "dumbing down" or "talking down" to the reader. (Office of Personal Management, 2017).

Pour promouvoir le style clair, les institutions s'appuient donc également sur le second type d'acteurs lié au style clair : les associations. On distinguera ici deux types d'associations.

Il y a tout d'abord des organismes qui offrent des services ((Balmford 2002)) et ressources pour aider à la rédaction en style clair, et militent auprès des acteurs institutionnels pour qu'ils l'utilisent. Pour cela, elles délivrent des récompenses ou apposent une "marque" aux documents écrits ou réécrits en style clair par les institutions juridiques et administratives, sous différents noms, par exemple (*Clear Mark*). Pour promouvoir le style clair, l'association Plain English Campaign (PLC) organise également des Plain Language Awards, afin d'encourager les sites webs institutionnels à utiliser le style clair à différentes échelles. À l'inverse, et non sans humour, l'organisme décerne également chaque année des récompenses aux journaux, personnalités politiques qui ne pratiquent pas la rédaction en *plain language*.¹⁴ Actives sur les réseaux sociaux, ces organisations publient fréquemment des recommandations et des exemples de textes en style clair en lien avec l'actualité. Nous présentons ici ce type d'acteur, car leurs marques nous ont aidé à concevoir notre corpus (cf. Chapitre 5).

L'autre type d'association est celui des organisations non-gouvernementales qui prennent en charge la diffusion du droit. Au Royaume-Uni, le Citizens Advice Bureau, une organisation britannique d'accompagnement juridique et administratif implantée à l'échelle nationale et locale, produit également des textes explicatifs de la loi, des droits et démarches pour les justiciables. tous les sous-domaines du droit et démarches juridiques et administratives, Ouverts en 1939 , es bureaux, peu à peu adossés aux autorités locales, accompagnent les usagers dans les grands mouvements de l'histoire législative et politique du Royaume-Uni. La version en ligne, qui existe depuis 1999, reflète également la numérisation croissante de la société.

Il est notable que la prise en charge de l'explication du droit est différente selon les contextes culturels ; il ne s'agit pas toujours du même interlocuteur. Si au Royaume-Uni, c'est donc une association, le CAB, qui est un acteur majeur de

14. Ce sont les *Gobbledygook Awards*, dont l'ancien président étasunien Donald Trump fut notamment l'un des récipiendaires en 2017.

l'accès au droit, en France, la majorité de la diffusion du droit est publiée par la Direction de l'Information Légale et Administrative (DILA), qui dépend du Premier Ministre, et s'adresse aux justiciables grâce à son site Service Public, source majeure de notre corpus francophone. Le Ministère de la Justice française, ainsi que les gouvernements locaux (départements) publient également des textes de diffusion du droit. Cette différence illustre des divergences culturelles entre les deux pays : prise en charge associative et culture *consumer-oriented* dans un cas, centralisation et concentration de l'information juridique dans l'autre. Au Canada, qui nous intéresse particulièrement en tant qu'Etat bilingue, c'est à la fois les institutions, comme la Cour Suprême, et les associations (comme par exemple l'organisme Educaloi) qui diffusent le droit auprès du grand public.

1.4.2 Fonctions et genres de la diffusion du droit

Tout comme le discours juridique renvoie à une multiplicité de situations discursives, la médiation juridique peut être décrite en termes de situations et d'objectifs communicationnels.

Puisqu'un nombre croissant de démarches juridiques et administratives s'effectuent en ligne¹⁵, nous nous concentrons ici sur le discours de diffusion du droit disponible en ligne, sous format numérique, bien que les caractéristiques globales soient similaires à la médiation papier. Les associations de promotion du style clair ont d'ailleurs su s'aligner sur cette numérisation mise en place par les pouvoirs juridiques et administratifs pour s'assurer que le contenu publié en ligne est compréhensible et lisible pour l'utilisateur, en octroyant des labels "ClearMark" ou "Plain Mark" à des textes en ligne. Cette évolution vers la numérisation, qui constitue un gain de temps pour les démarches administratives et certaines démarches judiciaires, tout en offrant un plus grand accès au droit par la mise à disposition de l'information juridique, peut également entraîner un accroissement des inégalités, par le passage à un nouveau support textuel numérique auquel tous les citoyens n'ont pas nécessairement accès à cause de fractures territoriales ou de problèmes d'accessibilité pour certaines populations (Défenseur des Droits 2019). Etudier le discours de médiation ou diffusion du droit présente donc des enjeux sociaux et politiques forts, qui invitent à interroger les fonctions et formes textuelles de cette diffusion.

Se pose tout d'abord la question de la catégorisation des genres web disponibles

15. Par exemple les demandes d'aide au logement, de papiers d'identité mais également plus récemment le dépôt de plainte pour certains délits.

pour le justiciable, question qui a été prise en charge par la linguistique de corpus suite à l'accessibilité accrue à un grand nombre de textes remplissant des fonctions diverses avec l'avènement du Web (Biber et Kurjian 2007). Avant de nous pencher sur les média et genres disponibles pour la diffusion du droit, il faut s'interroger sur la fonction de cette diffusion.

D'une part, certains textes ou documents ont pour fonction de faire connaître le domaine spécialisé, son fonctionnement, sa terminologie et ses derniers développements aux non-spécialistes de manière à ce qu'ils comprennent leurs droits, avec pour objectif de garantir et d'améliorer l'accès au droit. Les résumés des jugements publiés par la CSC, par exemple, remplissent cette fonction puisqu'ils ont pour but de permettre aux citoyen.nes de connaître les décisions prises par la plus haute instance judiciaire et qui ont des conséquences sur leurs propres droits :

Préparés par le personnel des communications de la Cour, ces résumés sont rédigés dans un langage accessible, afin de permettre à toutes les personnes qui le désirent de se renseigner sur des décisions qui ont une incidence sur elles. (Cour Suprême du Canada 2022)

Il s'agit donc ici d'un transfert de connaissances spécialisées, semblable en cela aux fonctions de la médiation scientifique. Une seconde fonction de la médiation juridique est de diffuser les connaissances spécialisées pour avoir un effet sur la prise de décision des usagers vis-à-vis d'une situation juridique ou administrative précise :

Ces pages ne se limitent pas à vulgariser un savoir disciplinaire théorique, mais transmettent un **pratique et procédural** sur ce qu'il convient de faire et quel comportement adopter en cas de contact avec la justice (Preite 2018).

Le terme "procédural" est ici crucial, puisqu'il fait entrer ces textes dans la catégorie des textes procéduraux ou d'incitation à l'action définis par Adam :

[Ces textes] sont destinés à faciliter et à guider la réalisation d'une tâche ou macroaction du sujet qui le souhaite ou est chargé de l'accomplir. (Adam 2011)

Adam inclut dans ce type de textes les documents ayant une fonction socio-pragmatique, allant des conseils publiés dans la presse à l'injonction ("Défense de fumer"), en passant par les textes "plus purement procéduraux" comme les recettes de cuisine, les modes d'emploi et les fiches de bricolage. Ces textes qui "disent de faire et comment faire en prédisant un résultat et en incitant très directement à l'action." (Adam 2001 : 11-12) . Les textes de médiation juridique correspondent en partie à ce type de textes, puisqu'ils expliquent aux justiciables, étape par étape, le

déroulement d'une procédure administrative ou juridique afin de les aider à obtenir un résultat. Ci-dessous sont présentés quelques exemples de titres de textes à fonction procédurale, qui viennent aider les justiciables à agir vis-à-vis d'une situation, par exemple de discrimination ou de harcèlement, à obtenir une allocation ou à accomplir une démarche :

- *Taking action about harassment*
- *Showing you're disabled under the Equality Act*
- *Fostering Handbook*¹⁶

Les deux fonctions principales dégagées sont donc des textes majoritairement informationnels et des textes majoritairement procéduraux (ou instructionnels), qui correspondent également aux catégories de Biber et al. (2015) pour classer les textes du Corpus of Online Registers of English. Les organismes qui promeuvent le style clair incitent donc les rédacteur.ices à s'interroger sur la finalité du texte, à savoir s'il s'agit plutôt d'informer les lecteurs, ou plutôt de les amener à prendre une décision ou encore les inciter à participer à un programme d'aide (Bureau de la Traduction, 2015). Cependant, beaucoup de textes ne présentent pas qu'une seule et même fonction, et doivent plutôt être qualifiés d'hybrides, car ils contiennent à la fois un volet informationnel et un volet procédural.

Dans tous les cas, ces textes, qu'ils soient informationnels ou procéduraux, présentent des enjeux différents que ceux de la diffusion du discours spécialisé scientifique, par exemple. Turnbull propose d'ailleurs de différencier ainsi la vulgarisation (*popularization*) de la médiation, qui n'est pas selon elle un simple transfert de connaissances :

This is where the main difference lies between popularizations, as for example on scientific discoveries, which aim to transfer information to broaden the readers' general knowledge, but probably will have no direct effect on his life, and mediation where information, whether it be legal, medical or otherwise, will actually be used by the reader to empower him/herself. (Turnbull 2018)

La médiation du discours juridique comporterait donc d'après Turnbull un enjeu social plus fort¹⁷, celui de combler l'écart entre le discours juridique, dont la complexité est perçue comme une violence institutionnelle, et les usager.ères du droit, en leur apportant un "savoir-que-faire" (Preite 2016) leur permettant de reprendre

16. Un guide destiné aux personnes souhaitant devenir famille d'accueil pour enfants placés.

17. Loin de nous l'idée de sous-estimer l'importance de la médiation scientifique en termes d'enjeux sociaux et politiques, mais plutôt de distinguer le transfert de connaissances théoriques et pratiques.

une forme de contrôle sur des situations juridiques et administratives à travers la connaissance de leurs droits.

Sous quelle forme et quels médias cette recontextualisation des connaissances juridiques se fait-elle ? Il existe un foisonnement des genres et des médias disponibles pour la diffusion des discours spécialisés en ligne, pas forcément juridique. Il existe des brochures et guides juridiques, documents de plusieurs dizaines de pages, qui ont souvent une fonction à la fois informationnelle et procédurale visant à guider une situation particulière (divorce, nomination en tant que juré de procès, demande d'asile); des fiches procédurales, courts textes des institutions expliquant une démarche démarches à suivre, qui diffèrent d'autres genres webs comme l'article de blog qui contient une date et une source énonciative définie; et les textes expliquant le rôle d'une institution, comme les résumés des décisions de la Cour Suprême du Canada. Ces genres, qui s'appuient sur un discours juridique originel spécialisé, remplissent donc une fonction informationnelle ou procédurale (ou les deux). Si les institutions et organisations qui les publient utilisent ces supports textuels, on trouve aussi des genres multimodaux pour diffuser le droit, comme par exemple des vidéos et infographies qui s'appuient sur des supports audio et visuels.

La diffusion du droit, comme la médiation scientifique, est donc caractérisée par une hybridité de modes et de genres résultant de la contamination de différents genres et média (articles, brochures, weblogs, vidéos) et répondant à différents objectifs, comme l'information ou l'instruction, mais également la promotion des activités de communication des institutions (Compagnone 2015). L'utilisation plus ou moins fréquente d'éléments multimodaux (comme les images) dans les textes procéduraux dépend également du lectorat visé.

1.4.3 Hétérogénéité des lectorats visés

Il faut souligner que, tout comme le lectorat du discours juridique est divers, Parmi les non-spécialistes, le lectorat n'est pas un ensemble homogène. C'est ce que soulignent les recommandations du Bureau de la Traduction du Canada, qui invite les rédacteur.ices à se demander si le lectorat visé est composé de "spécialistes, d'adolescents, du grand public ou de personnes pour qui le français n'est pas la langue maternelle ?" (Bureau de la Traduction du Canada, 2015). De même, Petelin (2010), dans ces recommandations pour rédiger en style clair, les incite à un travail précis d'identification :

Create a profile of intended reader : socio-eco background, age range, gender, ethnicity, geo location, educational background, familiarity with the subject, professional interests, sensitivities and motivations. (Petelin 2010)

Les recommandations pour les rédacteurs s’adressant aux justiciables insistent d’ailleurs sur la réflexion préalable nécessaire pour prendre en compte non seulement la variabilité sociologique du lectorat, mais également l’âge ou le niveau d’alphabétisation. Il existe en fait un continuum de la simplification linguistique selon les caractéristiques du lectorat. En particulier, certains textes sont rédigés pour un lectorat constitué de personnes présentant une déficience cognitive. Ce style, qui constitue un pas supplémentaire dans la simplification par rapport au PL, est appelé *Easy Language easy-to-read* et en français de Facile à Lire et Comprendre (FALC). Les textes rédigés en FALC font davantage usage de la multimodalité (images), tout en utilisant des phrases simples, contenant une seule proposition et aucun terme spécialisé, comme on le voit dans la Figure 1.3, qui montre la version FALC de l’attestation de déplacement dérogatoire publiée par le Ministère de l’Intérieur en 2020 en France lors de la crise sanitaire liée à la Covid-19.

FIGURE 1.3 – Exemple de document en FALC

**Je remplis cette attestation à chaque fois que je sors.
C'est la loi pour le coronavirus.
Les policiers peuvent me la demander.
Si je n'ai pas d'imprimante, je la recopie sur une feuille.**




 Je m'appelle :

 Je suis né le :

 Mon adresse :

Je me déplace exceptionnellement aujourd'hui :

  Pour aller travailler.

  Pour acheter à manger.

  Pour aller chez le docteur.
Pour aller à la pharmacie.

  Pour m'occuper d'une personne
qui a besoin de moi.

  Pour sortir peu de temps pas loin de chez moi.
Pour sortir mon chien.

 Je mets la date d'aujourd'hui :

 Je mets l'heure de ma sortie :

 Je signe :

Originellement développé par (et pour) les personnes présentant des troubles des fonctions cognitives, le FALC et l'*Easy Language* sont également utilisés pour la communication avec les personnes allophones ou peu à l'aise avec la lecture (Gangloff 2015). Il existe donc différents degrés de simplification selon le lectorat visé et différentes stratégies selon l'âge ou les spécificités. La littérature sur la diffusion des connaissances juridiques ou administratives a pu porter sur des segments spécifiques de la population, notamment comment était reformulés le discours du droit européen pour les enfants (Diani 2018), les conditions générales de vente de produits financiers pour les personnes âgées (Rossetti et al. 2020), ou encore le droit pour des étudiant.e.s en droit suivant un cours d'anglais spécialisé L2 (Dąbrowski 2016).

Les recommandations pour un style clair portent en général sur un lectorat ne présentant pas de handicap cognitif, c'est pourquoi ils n'ont pas été inclus dans cette étude. En effet, les textes auxquels nous nous intéressons s'adressent (en général) à des adultes, dont le niveau d'alphabétisation est considéré comme "moyen" (cette notion de lectorat moyen et de "grand public", nous le verrons, n'est pas sans poser problème dans les mesures de complexité et de lisibilité d'un texte).

1.4.4 Stratégies de reformulation et d'explicitation des discours spécialisés

Le discours juridique doit être adapté au lectorat non-expert par une reformulation du contenu spécialisé. Les études sur la médiation s'accordent à dire que passe par une recontextualisation des connaissances spécialisées :

The rewriting enables the reader to see the situation from his experiential point of view, so that abstract knowledge is associated with familiar situations and becomes practical knowledge. (Turnbull 2018)

Cette reformulation se fait selon des stratégies transversales aux discours spécialisés. Dans le Tableau 1.4 est reproduit l'exemple (a) du Tableau 1.2 ainsi que sa reformulation en style clair.

Turnbull (2018) distingue deux dimensions pour le transfert des connaissances spécialisées. Tout d'abord, la dimension cognitive, qui prend la forme d'une concrétisation (*concretization*), c'est-à-dire toutes les stratégies de reformulation d'information abstraite faite de manière à relier la réalité de l'expert.e à celle du non-expert.e en mettant en rapport l'information spécialisée et des situations familières pour les non-expert.es ou grâce à des concepts alternatifs (Turnbull 2018). Ces stratégies de

TABLEAU 1.4 – Exemple de reformulation du discours juridique (législatif)

EN-Lois	EN-CAB
(D1) A person A discriminates against a woman <u>if, in the period of 26 weeks</u> beginning with the day on <u>which she gives birth</u> , A treats her unfavourably <u>because</u> she has given birth.	(D2) Pregnancy and maternity discrimination is when you 're treated unfairly because you're pregnant, breastfeeding or because you've <i>recently</i> given birth.

concrétisation incluent la description, la définition, comme on le voit dans l'exemple (D2) du Tableau 1.4, qui illustre définition de *pregnancy discrimination*, mais aussi l'exemplification, avec des exemples concrets d'éléments, phénomènes ou situations. Sous forme plus élaborée, une de ses stratégies est les scénarios, qui présentent des situations plus complexes, comme on le voit ci-dessous dans l'exemple (18).

- (18) Example : You're breastfeeding your 6-week-old baby on the bus. The driver tells you to stop breastfeeding or leave the bus, because another passenger has complained. This would be unlawful pregnancy and maternity discrimination, as you're within the 26 weeks time limit (EN-CAB).

La dimension cognitive passe également par la paraphrase de concepts, et la répétition dans le but d'anticiper les difficultés et efforts cognitifs nécessaires pour les lecteurs et lectrices.

La seconde dimension du transfert de connaissances spécialisées est la dimension communicative, c'est-à-dire les stratégies communicatives qui passent par la personnalisation. Comme vu plus haut, le discours juridique est hautement abstrait et impersonnel (Person A). La reformulation va donc passer en partie par ce que les auteurs de *Critical Discourse analysis* appellent *conversationalization of public discourse* (Turnbull 2018), notamment avec l'utilisation de la deuxième personne du singulier *you* mais également les formes contractées orales, pour mettre en place une forme de dialogisme avec le lectorat non-spécialiste. La dimension communicative prend aussi la forme de marqueurs de politesse en particulier en anglais. Les conseils, recommandations et l'incitation à l'action par des experts peuvent en effet être considérés comme des actes menaçants pour la face (*face-threatening acts* en contexte anglo-saxons, notamment parce qu'ils peuvent apparaître comme intrusifs ou comme une menace pour l'autonomie des personnes qui les reçoivent (MacGeorge et Van Swol 2019). Cette question a été partiellement explorée par Dieckmann et Rojas-Lizana (2016) dans le cadre de la communication entre avocat.es bénévoles

et client.es lors de permanences juridique en Australie, dans une étude portant sur les stratégies de solidarité et de politesse mise en place par les expert.es pour minimiser le déséquilibre lié à leur statut. La structuration des textes de médiation se fait également par le biais de phrases interrogatives (Turnbull 2018), qui servent à anticiper les questions que peuvent se poser les lecteur et lectrices, comme on le voit dans les extraits de textes destinés au grand public ci-dessous.

- (19) What can I expect at my asylum interview? (...) What happens after my asylum interview?
- (20) Quelles sont les sanctions si [je suis nommé jury et] je ne viens pas? Puis-je être dispensé ou excusé?

Une autre stratégie importante repose sur la multimodalité, notamment la mise en page ou l'utilisation d'images, diagrammes et figures pour illustrer une procédure, par exemple. Nous venons donc de voir globalement les stratégies communicative et cognitive employée pour passer d'un discours destiné principalement à un lectorat spécialisé à un discours destiné s'adressant lectorat non expert. Le *plain language* a notamment été développé pour faciliter et guider cette reformulation, comme nous allons le voir.

1.5 Résumé du Chapitre 1

Dans ce chapitre, nous avons introduit le domaine du droit, caractérisé par une grande hétérogénéité des genres et situations discursives, tout comme l'est celui de la médiation juridique. Les caractéristiques de l'anglais (et du français) juridique ont ensuite été présentées, notamment sa terminologie hautement spécialisée, souvent latinisante, son haut degré d'abstraction, et sa syntaxe complexe, auxquelles s'ajoutent les noms prédicatifs et les tournures impersonnelles (Williams 2004). Ces caractéristiques répondent à plusieurs besoins parfois contradictoires de généricité et de précision, de formalité mais également d'un besoin de rester vague. Nous avons ensuite évoqué les différentes institutions et associations qui prennent en charge la diffusion des discours législatif et judiciaire dans les Etats qui nous intéressent pour la constitution de notre corpus : Royaume-Uni, Nouvelle-Zélande, Canada et France. Plusieurs stratégies discursives et rhétoriques permettent de reformuler les connaissances spécialisés du droit. Dans le prochain chapitre, nous nous attelons à une défi-

dition et à une description plus précise des caractéristiques du PL, qui a été développé comme l'une de ces stratégies linguistiques pour rendre les discours spécialisés accessibles aux non-spécialistes.

Chapitre 2

Le style clair en droit : histoire, définitions et enjeux linguistiques

Le *plain language*, né dans la seconde moitié du XXI^{ème} siècle sous l'impulsion du Plain Language Movement (PLM), s'appuie sur une série de recommandations et conventions linguistiques à l'intention de rédacteur.ices, traducteur.ices et institutions pour encourager une communication claire avec leur lectorat-cible. Notre objectif est de comparer en corpus des textes du discours juridique réputés complexes à leur version proclamée "*plain language*". Dans ce chapitre, nous proposons un état de la question de l'histoire de ce concept, notamment en droit, ainsi qu'une discussion de ces définitions et surtout de ces caractéristiques linguistiques. Parce qu'il est sous-tendu de puissants enjeux sociaux et politiques, le style clair - à la fois concept linguistique, service et objet de revendications - est en effet un phénomène multiforme. Nous présentons ensuite brièvement quelques exemples de textes juridiques reformulés.

2.1 Le style clair : définitions et champs d'application

Il est tout d'abord nécessaire, avant de caractériser linguistiquement le PL, de le replacer dans le cadre linguistique plus large des langues planifiées et des langues simplifiées.

2.1.1 *Plain language* et autres langues simplifiées

Le PL est défini comme suit par les associations et organismes qui promeuvent son utilisation, en particulier dans le cadre du pour le discours du droit :

A communication is in plain language if its language, structure, and design are so clear that the intended audience can find what they need, understand what they find, and use that information. (Plain Language Association International 2022)

Selon cette définition, qui souligne l'idée d'une intervention consciente et planifiée sur une langue naturelle, le PL, peut être rattaché plus généralement au groupe des langues planifiées et est lié plus globalement aux concepts de planification et de politique linguistique. Les langues planifiées incluent des Langues Auxiliaires Internationales comme l'esperanto, dont l'objectif affiché lors de sa création était de proposer une langue universelle, restreinte et plus aisée à apprendre que les langues naturelles, et transcendant les frontières (Gobbo 2017). Les langues inventées ou imaginaires créées pour des œuvres de fiction (elfique, klingon) appartiennent également aux langues planifiées. Enfin, on trouve dans cette catégorie les langues contrôlées, aussi appelée langues simplifiées, qui ont un lexique et un nombre de structures grammaticales restreints, et sont définies comme un sous-ensemble d'une langue naturelle, construite pour la rédaction ou la communication dans un contexte spécifique :

A controlled language is a constructed sub-variety of language (often, but not always English), which has been specifically designed for professional communication, usually by a large commercial or governmental organisation. (Gledhill 2011)

On peut déjà noter ici que le PL diverge légèrement de cette définition, puisqu'il n'est pas conçu uniquement comme le support d'une activité technique ou pour la communication entre professionnels d'un domaine, mais comme un moyen de communiquer avec les non-spécialistes. Les langues contrôlées (LC) sont donc des sous-ensembles d'une langue naturelle (l'anglais, mais aussi le français, l'allemand ou encore le suédois) dont le lexique, la grammaire et la syntaxe ont été restreints pour un objectif communicatif précis, qui est généralement de réduire la complexité ou l'ambiguïté. Elles se présentent souvent sous la forme de recommandations qui portent sur la syntaxe, la grammaire, la longueur des phrases, les termes et expressions à employer (Gledhill 2011). Le *Basic English*, l'une des plus célèbres langues planifiées, fondée par le sémanticien et philosophe Ogden et Halász (1935) visait, avec son vocabulaire de 850 mots seulement, à être rapidement et facilement assimilable pour les apprenants allophones de l'anglais. Le *Basic English* a inspiré la

création de plusieurs langues simplifiées. A ce jour, un pan entier de l'encyclopédie collaborative digitale Wikipedia est rédigé en *Simple English*, dérivé du *Basic English*, et a été utilisé comme corpus notamment dans les recherches en Traitement Automatique du Langage (TAL) pour explorer la simplification automatique de phrases (Vajjala et Meurers 2014). D'autres langues simplifiées ont vu le jour dans des domaines spécialisés, afin de rendre plus efficace la communication entre professionnels. C'est le cas du *Simplified Technical English* (STE), dérivée du Basic English d'Ogden, une langue contrôlée internationale standardisée, d'abord développée par l'aérospatiale étatsunienne. Le STE vise en effet une clarté semblable à celle du *plain language* : Il est utilisé et étudié particulièrement dans le cadre de la rédaction technique. Souvent obligatoire pour la rédaction de manuels et de documentation technique, en particulier liées à la maintenance, dans différentes entreprises (Airbus, Boeing) et organisations militaires ou gouvernementales, le STE a fait l'objet d'une standardisation internationale et a entraîné la création d'outils qui permettent aux organismes l'utilisant de vérifier automatiquement que leur documentation est bien conforme aux normes lexicales, syntaxiques et textuelles du STE : ces outils sont liés à la notion de conformité et de respect des règles rédactionnelles (*compliance*), que nous allons pour notre part explorer dans le cas du *plain language*. Plusieurs langues contrôlées ont également été créées dans les années 1960-1970, dans des domaines d'application variés de la rédaction et de la traduction techniques, notamment dans les cadres militaire, gouvernemental et industriel. Cependant, toutes les langues simplifiées ne sont pas créées ou fondées sur la langue écrite ; certaines sont au contraire développées pour la communication orale. C'est le cas de l'*ATC phraseology*, une série de phrases et mots standardisés créée pour les interactions entre contrôleurs aériens et pilotes. Cette langue contrôlée a notamment été mis en place à l'échelle internationale pour diminuer le risque d'ambiguïté et d'incompréhension entre ces acteurs du domaine aéronautique, dans des situations communicationnelles à haut risque où des erreurs de compréhension peuvent mener à des accidents parfois mortels (Lopez et al. 2013).

Le *plain language*, qui a pour but de reformuler des textes juridiques complexes, est donc un concept proche des langues simplifiées, puisqu'il repose sur des recommandations qui restreignent le lexique et certains traits grammaticaux. On notera cependant que les recommandations rédactionnelles pour le PL sont considérées comme de simples directives. Petelin (2010) suggère que c'est la visée communicationnelle et la clarté, plutôt que le respect des règles, qui font le style clair :

Writing in plain language is just writing in clear, straightforward language, with the needs of the reader foremost in mind. There are no hard-and-fast rules. The main

thing to remember is that if what you have written could be unclear or confusing for your reader, or difficult to read, you should rewrite it so that it becomes clear, unambiguous and easy to read. (Petelin 2010)

De même, pour Eagleson, le style clair n'est pas une simplification de l'anglais comme le STE, mais un usage juste et précis de la langue :

Plain English is clear, straightforward expression, using only as many words as are necessary. It is language that avoids obscurity, inflated vocabulary and convoluted sentence construction. It is not baby talk, **nor is it a simplified version of the English language**. Writers of plain English let their audience concentrate on the message instead of being distracted by complicated language. They make sure that their audience understands the message easily. (Eagleson et al. 1990)

Le *plain language* est donc davantage un mode discursif qu'une langue contrôlée ; c'est pourquoi nous le traduisons par "style clair". Toutefois, ce concept est traversé par des enjeux communs aux langues contrôlées : la notion récurrente de règles pose, dans les deux cas, la question du prescriptivisme et du degré d'application réel de ces recommandations.

2.1.2 Le style clair, un principe de rédaction qui touche à différents domaines

Le PL est un concept qui touche à différents domaines, puisqu'il s'agit d'un principe stylistique qui peut être appliqué à différents contextes. Il faut tout d'abord noter que la plupart des articles de recherche sur la question du PL, de ses enjeux, de sa mise en pratique et de sa réception sont publiés dans des revues spécialisées, notamment des revues de médecine, de communication de la santé ou de droit. Ainsi l'article liminaire de Charrow et Charrow (1979), qui relate une expérience de psycholinguistique sur les instructions écrites données aux jurés aux Etats-Unis, fut-il publié dans une revue de droit. Les chercheurs en psycholinguistique et en psychologie cognitive se sont ensuite intéressés à la réception des textes estampillés PL. Par exemple, certains auteurs ont étudié la réception de textes après différentes phases de réécriture, afin de voir les différences de temps de lecture pour les participants avant et après simplification, notamment de contrats bancaires ou d'assurance (Masson et Waldron 1994). Cependant, le style clair n'est pas utilisée uniquement dans le domaine juridique et administratif. Tout d'abord, le domaine des affaires a également été visé par le Plain Language Movement, notamment pour la simplification des

documents bancaires à valeur contractuelle (Asprey 2004). L'importance de l'utilisation du style clair pour le domaine commercial et financier est soulignée par Bruce Bradley, le député étasunien qui a introduit le *Plain Writing Act* au Congrès des Etats-Unis : il souhaitait rendre lisibles les régulations fédérales pour les petites et moyennes entreprises¹. L'impact du PLM sur les textes du domaine financier a donc été étudié pour mesurer le degré de respect des guides rédactionnels, par exemple par Loughran & MacDonald (1999). Ces derniers ont étudié les conséquences de la *Plain English rule* de la Security and Exchange Commission en 1998². Leur étude porte sur un genre spécifique, celui des prospectus destinés aux investisseurs et détaillant la situation financière de l'établissement concerné. Le langage clair et simple est donc utilisé à l'intérieur du milieu professionnel des affaires, dans le but de simplifier la communication entre spécialistes et de les inciter à investir grâce à une communication efficace.

Outre les domaines du droit, de l'administration et des affaires, le domaine médical est également actif dans la production de textes en PL. La médecine est par ailleurs l'un des domaines principaux où se pose la question de la clarté rédactionnelle (mais pas nécessairement du *plain language*). Nombreuses sont les études de lisibilité sur différents genres de diffusion ou de vulgarisation des connaissances spécialisées, en ligne ou sur papier, en particulier les brochures ayant un objectif d'éducation thérapeutique pour un lectorat de malades adultes ou adolescents (Nagel et al. 2008). Des organisations non gouvernementales se réclament du PL, notamment l'organisme Cochrane, qui propose des revues de la littérature médicale sur certaines procédures, traitements, maladies, affections ou syndromes précis. En parallèles de ces synthèses, qui sont destinées à des spécialistes, l'organisation diffuse des *plain language summaries*, c'est-à-dire des résumés, rédigés selon les principes du style clair et destinés à un lectorat divers. Cochrane étant une organisation internationale présente dans plusieurs pays, ces résumés sont d'abord rédigés en anglais avant d'être traduits, partiellement de manière automatique. Les résumés en style clair publiés par Cochrane sont donc un objet de recherche particulièrement intéressant pour étudier les stratégies de reformulation de ce discours spécialisé. Gledhill et al. (2019) ont par exemple étudié les caractéristiques textométriques et phraséologiques des

1. Gobbledygook dominates the regulations issued by government agencies, making it almost impossible for small businesses to understand the rules of the road. (Press Release, Office of United States Congressman Bill Bradley, 2012. *Bradley Introduces Bill to Streamline Regulations, Save Small Businesses Money*).

2. "In October 1998, the SEC implemented a rule requiring firms to use plain English in their prospectus filings. (...) The impetus of the rule is that investors will be better able to assess and more likely to invest in companies whose financial disclosures are not buried in legal jargon and obtuse language."

résumés en PL, pour identifier les stratégies utilisées pour simplifier, reformuler et clarifier le discours médical à partir des revues systématiques Cochrane. Jelacic Kadic et al. (2016) ont examiné le respect ou le non-respect des feuilles de style de l'organisme Cochrane et ont constaté qu'elles sont peu respectées, notamment en ce qui concerne la macro-structure. Ces études sont proches des problématiques qui nous concernent pour le domaine du droit : nous nous intéressons également à la question de la conformité des textes se réclamant du style clair.

Enfin, le style clair a plus récemment été mis en avant dans le domaine de la recherche scientifique. Certain.e.s ont appelé les auteurs d'articles de recherche et livres scientifiques à utiliser le PL également, afin de combattre la réputation d'un manque de transparence dans la recherche auprès des publics non-experts. Certains sites Web se spécialisent même dans la production de tels textes. C'est le cas par exemple de la plateforme en ligne *Brainstorm*, qui a vu le jour à l'initiative de la chaîne de radio et de télévision publique irlandaise *Raidió Teilifís Éireann*, et qui publie des articles rédigés en PL, afin de diffuser largement les dernières recherches dans leurs domaines³.

L'intérêt porté par les chercheurs pour le style clair est également visible par le nombre de colloques et conférences internationaux (plus ou moins) récents, notamment la conférence PLAIN, tenue à Montréal en 2018. C'est notamment grâce à ces conférences que certains genres textuels de notre corpus ont pu être choisis. Si la littérature regorge donc d'études et de réflexions sur la LCS dans différents domaines (droit, médecine, milieu des affaires, recherche) et selon différentes approches (psycholinguistique, traitement automatique des langues, traduction), rares sont les études qui ont comparé plusieurs genres discursifs du droit avec leur version claire et simple en anglais et en français de manière systématique.

2.2 Le style clair en droit : histoire et développement

Dans cette section, nous survolons l'histoire du *Plain Language Movement* dans les pays et institutions où il s'est développé et étendu dans la seconde moitié du vingtième siècle, avant de voir la situation de la langue claire et simple dans le

3. 'Brainstorm : Where academics write interesting articles in plain English', Carroll, May 26 2018. Online. URL : <https://www.irishtimes.com/culture/tv-radio-web/brainstorm-where-academics-write-interesting-articles-in-plain-english-1.3504486>. Consulté le 9/4/2018

monde, en particulier dans les pays francophones.

2.2.1 Développement historique du *plain language* en droit dans les pays anglo-saxons

Si le cœur du Plain Language Movement s'est développé sous ce nom dans la deuxième moitié du 20^{ème} siècle, la critique des langues spécialisées, notamment juridiques, dépeintes comme contenant trop de "jargon" et comme inutilement complexes remonte à plusieurs siècles. Sir Matthew Baillie Begbie, juge en chef en Colombie Britannique de 1870 à 1894, écrit par exemple :

The Statute books are exceedingly muddled. I seldom look into them (cité dans Masson et Waldron (1994))

Cutts (2009) remonte quant à lui jusqu'à Edward VI, qui se plaignait de la complexité de la langue juridique et refusait de lire les textes de lois. Plus tard, des ouvrages dénoncent une langue bureaucratique et technocratique, comme *Gobbledygook Has Gotta Go* publié par O'Hayre (1923). Le terme de *gobbledygook* est d'ailleurs souvent employé par les auteurs qui jugent les langues spécialisées incompréhensibles, et défini comme "wordy and generally unintelligible jargon" (Merriam-Webster 2015) (nous reparlerons de ce mot en lien avec les mesures de complexité linguistiques dans le Chapitre 3). Aujourd'hui encore, la culture populaire fait régulièrement la satire du langage juridique. Ces exemples, qui peuvent paraître anecdotiques, montrent que la perception des langues spécialisées comme obscures et complexes est ancrée dans l'imaginaire collectif. Nous l'avons vu dans le Chapitre 1, le droit est caractérisé par une forte asymétrie de la relation entre expert.es et non-expert.es. Les défenseur.euses du style clair soulignent la dimension oppressive que renferme un texte rédigé en langue spécialisée pour le grand public :

[The texts] provide information that, if misunderstood or half understood, disadvantages people, oppresses them, or, - at the least - wastes their time and money. (Cutts 2009 : x).

C'est cette asymétrie entre experts et non-experts, en particulier juridiques ou bureaucratiques, qui est à l'origine du *Plain Language Movement*. Celui-ci naît dans différents pays anglophones dans les années 1970, avec l'avènement de groupes de consommateurs.ices, qui réclament davantage de transparence de la part des institutions administratives et juridiques. L'objectif de ce mouvement est de lutter

contre des discours jugés incompréhensibles pour les citoyens ou consommateurs, en promouvant l'utilisation d'une langue claire et simple dans les documents destinés au grand public. Le PLM revendique le droit pour le citoyen, consommateur ou patient destinataire du texte, d'avoir pleine connaissance de ses droits ou de prendre une décision éclairée pour sa santé : il s'agit des enjeux d'accès au droit mentionnés au Chapitre 1.

Les pays pionniers en termes de langue claire et simple sont les Etats-Unis et le Royaume-Uni, mais également l'Australie et la Nouvelle-Zélande, où les associations de consommateurs et citoyens formèrent un mouvement populaire réclamant la simplification des documents administratifs et juridiques, en particulier des contrats issus du droit privé. Si l'un des premiers documents publiés en style clair provient d'une banque, la First National City Bank (désormais Citibank), qui publie en 1975 une note sur les prêts bancaires rédigée en PL, les institutions publiques étasuniennes s'engagent rapidement en faveur du style clair, et plusieurs décrets présidentiels encouragent l'utilisation d'une langue claire et simple dans les textes et documents destinés au grand public (Petelin 2010). Le mouvement pour la simplification des documents de communication publique aux Etats-Unis culmine avec la proposition de loi du député Bruce Bradley intitulée *Plain Language Bill*⁴. La loi résultant de cette proposition, le *Plain Writing Act*, est votée et promulguée en 2010 par le président Barack Obama et stipule que les documents publiés ou transmis par les agences fédérales doivent être rédigés ou réécrits en LCS. Les directeurs d'agences fédérales doivent par ailleurs publier des rapports expliquant comment ils comptent appliquer la nouvelle loi, et nommer des agents en charge de l'exécuter. Si les textes composant nos corpus ne sont pas issus des Etats-Unis, cet exemple permet de faire apparaître ici en creux l'une des questions qui, nous l'avons vu, sous-tend l'étude du style clair : quel est le degré d'application réel des principes du style clair dans les documents produits ? Comment contrôler l'application de ces principes dans des institutions qui produisent chaque jour des milliers de textes ? On notera par ailleurs que la loi de 2010 sur le style clair aux Etats-Unis ne concerne pas le discours législatif, c'est-à-dire les textes de loi, mais les documents émis par les agences fédérales à l'intention du grand public.

Au Royaume-Uni, le *Plain Language Movement* prend son essor au milieu des années 1970, notamment après la publication du rapport de la Commission Renton de 1975, chargée d'étudier le style juridique anglais et de proposer des recommanda-

4. "[The bill] require[s] the federal government to write documents such as tax returns, grant application forms, web sites, and pension forms in simple, easy-to-understand langage [in any document other than a regulation]".

tions pour le simplifier (Taylor 2003). En parallèle, des associations comme la *Plain English Campaign* et des experts-linguistes activistes, comme le Cutts, militent à partir de 1979 contre le "jargon" et le "charabia". Depuis, de nombreux efforts ont été faits par les institutions, d'une part pour simplifier non seulement la rédaction des contrats (d'assurance notamment) et des textes législatifs (Williams 2004, Richard 2006), mais également pour rendre claire et simple la communication entre les institutions et les justiciables. Le site gouvernemental *gov.uk* affirme ainsi utiliser une feuille de style conforme aux conventions du PL, qui est érigé en véritable principe communicationnel :

Plain English is mandatory for all of GOV.UK.(...) Plain English is the whole ethos of GOV.UK : it's a way of writing.⁵

Cela montre l'impact du PLM sur les politiques de communication des autorités publiques. Des associations comme le Citizens Advice Bureau, l'une des sources pour notre corpus, déclarent également utiliser le PL comme principe rédactionnel.

En Nouvelle-Zélande, également fer de lance du mouvement, le PLM a eu une incidence tout d'abord sur la rédaction de la législation, en particulier la réécriture en style clair des lois sur les impôts sur le revenu en 1996 (Asprey 2004 : 33). Suite à cela, un nouveau format de législation a vu le jour en 2000, afin de réécrire ou d'écrire les nouvelles lois suivant un format clair, disponibles publiquement sous forme imprimée ou en ligne. Le Parliamentary Counsel's Office publie des recommandations détaillées non seulement pour s'adresser au grand public mais également pour réduire l'ambiguïté et la complexité dans les textes législatifs (Asprey 2004 : 33). Enfin, pour ce qui est de la communication avec le public, les rédacteur.ices du site du gouvernement néo-zélandais DIGITAL.GOV.NZ ont pour ligne éditoriale d'utiliser le PL (Digital Government New Zealand 2020).

Au Canada, les principes du style clair sont utilisés dans les textes législatifs (Office of the Scottish Parliamentary Counsel 2006), dans les formulaires publiés par les agences fédérales gouvernementales, ainsi que les contenus en ligne adressés au grand public (Asprey 2004 : 10). Le style clair est également incorporé à la diffusion du discours judiciaire et juridictionnel, au niveau des plus hautes instances juridiques du Canada. La Cour Suprême du Canada publie en effet des résumés jugements en style clair, comme l'expliquait le Juge président de la Cour Suprême Wagner en 2018 : "*You have to make sure there are clear decisions accessible in clear language.*"

5. "Content design : planning, writing and managing content", 10 March 2020 [Online] URL : <https://www.gov.uk/guidance/content-design/writing-for-gov-uk/plain-english>. Consulté le 31/03/2020

(Fine, 2018). Ce sont ces genres, similaires aux résumés des revues systématiques Cochrane, qui vont notamment nous intéresser.

Ainsi, dans les pays fondateurs du PLM, différentes instances juridiques se sont emparées de la question de la clarté rédactionnelle, à différentes échelles et pour différents publics. Le PLM a conduit à différentes mesures de politique linguistique prises par les autorités publiques dans le domaine du droit, en particulier dans les pays dont nous étudierons les productions, le Canada, la Nouvelle-Zélande et le Royaume-Uni. Qu'en est-il ailleurs dans le monde ?

2.2.2 Le *plain language* dans le monde

Le PLM a fait des émules dans d'autres institutions et pays du monde, pour d'autres langues, que nous évoquons brièvement. La question de la clarté rédactionnelle en droit, à la fois dans les textes adressés aux experts et ceux adressés au grand public, est en effet un enjeu pour les institutions internationales et les États multilingues. À l'échelle de l'Union Européenne, la question de la simplification d'une langue technocratique se retrouve aussi dans la campagne "Fight the FOG" menée en 1999 par les traducteurs officiels (Williams 2004). La Commission européenne publie également des recommandations pour encourager les rédacteurs et traducteurs à utiliser la langue claire et simple, sous le titre *How To Write Clearly* (Union Européenne, 2011), disponible dans les 23 langues officielles de l'Union Européenne. L'utilisation du PL se pose également dans certains autres pays du Commonwealth, en particulier pour les rédacteurs du discours législatif, héritiers du système juridique britannique et où l'anglais est souvent toujours une langue officielle, par exemple en Malaisie (Hashim 2013). À différentes échelle, donc, diverses institutions tentent de mettre en œuvre une politique de clarification de la langue juridique ou s'interrogent sur la clarté rédactionnelle en droit.

Nous avons pour l'instant mentionné principalement les pays et institutions de langue-culture anglo-saxonne qui utilisent le style clair, qu'il soit désigné *plain language* ou *plain English*. De nombreuses initiatives ont également vu le jour pour d'autres langues. La dimension multilingue de notre recherche ne concernant que le français et l'anglais, nous évoquons rapidement les initiatives dans d'autres langues. Williams (2004) évoque par exemples les langues claires et simples fondées sur le suédois et le finlandais, que les institutions publiques notamment utilisent pour clarifier le discours du droit. La revue *Clarity*, fondée par les traducteurs.ices de l'Union Européenne, recense les initiatives prises pour la simplification de documents juridiques

dans différentes langues, comme par exemple l'italien, l'espagnol ou le japonais (Asprey et Rawson 2005). En Allemagne, une langue similaire au FALC, destinée à faciliter la compréhension de textes pour les personnes présentant des déficiences cognitives a également vu le jour, par exemple pour simplifier des textes comme la Déclaration universelle des droits de l'homme (Bock 2018) : c'est la *Leichte Sprache*.

2.2.3 La clarté rédactionnelle en droit dans les pays francophones

Le besoin de clarification des langues spécialisées, notamment dans le domaine juridique, a également résulté en des mouvements de simplification dans les pays francophones. Au Canada, la notion de PL est traduite par "langue claire et simple", et, dans le cadre de la politique de bilinguisme, les principes du style clair sont également appliqués au français :

Canada is dual lingual and, although the Conventions originally applied only to the drafting of the English versions of legislation, they now extend to the French text also (Office of the Scottish Parliamentary Counsel 2006).

Le gouvernement du Québec et les avocats du barreau québécois se réclament également du *Plain Language Movement* et appliquent ses principes de rédaction. Plusieurs organismes, cabinets de rédactions juridiques ou associatifs, prennent en charge la diffusion du savoir juridique pour le Canada francophone. Lahlou (2013) recense les initiatives de clarification de la langue du droit dans différents pays francophones dans la seconde moitié du vingtième siècle, et mentionne l'association belge Droits Quotidiens, créée au milieu des années 90 à l'initiative de juristes de l'Université de Louvain, qui vise à donner de l'information juridique dans un langage clair. Cette association, constituée d'expert.es, vise des lectorats et des missions multiples. Ils s'adressent d'abord à des professionnels du droit, juristes et travailleurs sociaux.

En France, la question de la simplification de la langue du droit et du droit en général, en particulier des textes législatifs, est abordée régulièrement par les gouvernements et institutions, par exemple le Guide de légistique du Conseil d'État (Conseil d'État 2017). Plusieurs initiatives ont vu le jour pour s'adresser aux citoyens.ne.s dans une langue claire, en particulier dans les courriers administratifs⁶.

6. Le logiciel LARA, mis en ligne en 2002 par le gouvernement français, avait pour but de guider les rédacteurs de l'administration. (Clubic 2002)

Le Comité d’Orientation pour la Simplification du Langage Administratif (COSLA), en collaboration avec le Centre de Langues Appliquées de Besançon, a de plus publié un Guide Pratique de la Rédaction administrative (Colette et al. 2002), dans le but de simplifier les courriers envoyés par les institutions administratives. Comme nous le verrons, ce guide rédactionnel contient des recommandations similaires aux manuels de PL : on peut donc ici aussi parler de style clair, selon le sens établi en introduction. Une branche française de l’association belge Droits Quotidiens publie depuis peu des fiches juridiques explicatives concernant le droit français ; cette association s’appuie sur les notions de langage clair mais également sur de *legal design*, une approche transdisciplinaire du droit qui met en lien juristes et designers, afin de présenter des contenus juridiques de manière claire. On voit donc la diversité des acteurs et des approches de la clarté rédactionnelle en droit, notamment pour s’adresser aux justiciables, mais aussi parfois pour simplifier les discours spécialisés. Le style clair est un concept qui apparaît à différents degrés, dans des langues et pays de cultures juridiques différentes. Notre travail vise donc à étudier les productions se réclamant du style clair à différents niveaux, soit que le concept existe depuis longtemps, comme au Royaume-Uni, en Nouvelle-Zélande et au Canada, soit qu’il commence tout juste à être utilisé, comme en France. Les textes de nos corpus sont donc issus de ces quatre pays, car ils présentent différentes formes et degrés d’utilisation de la langue claire et simple, en particulier dans le cadre du droit et de sa médiation.

2.2.4 Le style clair comme service

Balmford (2002) affirme que le style clair n’est plus seulement un mouvement (au sens de mouvement politique), mais également un service offert par des entreprises et organisations. La notion présente donc des enjeux non seulement politiques et sociaux, mais aussi économiques. En effet, la sphère du *plain language* est devenue un marché, car des acteurs commerciaux proposent à la fois des réécritures de documents et des formations pour rédacteurs aux entreprises et institutions publiant des textes techniques. Les institutions publiques elles-mêmes offrent également des formations au Plain Language. Pour exemple, l’*Australian Public Service Commission* propose par exemple à ses employés des formations d’une journée ("*workshops*") leur permettant d’apprendre à utiliser, pour leurs propres écrits, les principes du style clair, incluant les emails, rapports et mémoires⁷. La concision et l’amélioration de

7. 'Plain English', Australian Public Service Commission, [Online] URL : <https://www.apsc.gov.au/plain-english>. Accessed 9/3/2018.

l'écriture de l'employé sont vantées comme les objectifs principaux de ce type de formation. L'apprentissage et l'utilisation du style clair sont en effet, pour les autorités, institutions et entreprises, un moyen de réduire les problèmes de compréhension des client.e.s ou justiciables, et ainsi de faire des économies en réduisant le temps d'appels. Côté francophone, des organismes de rédaction technique se sont également spécialisés dans les formations permettant d'apprendre à utiliser les principes du style clair, comme au Canada francophone, ou plus récemment en France⁸. Droits Quotidiens, l'association Belge, propose également des formations aux juristes.

Les services offerts par les organismes se réclamant du style clair ne se limitent pas à des réécritures ou à un travail de formation. Certaines organisations sont devenues des instances qui sanctionnent les institutions ou entreprises en leur attribuant des récompenses ou une marque, un tampon "clair et simple" qu'elles peuvent apposer sur leurs documents ou sites webs, comme on peut le voir avec la Crystal Mark de l'association Plain English Campaign :

The Crystal Mark is our seal of approval for the clarity of a document. Launched in 1990, and the first mark of its kind, the Crystal Mark is now used by over 2,000 organisations who want to provide the clearest possible information. (Plain English Campaign)

Au Royaume-Uni et en Nouvelle-Zélande, des organismes qui promeuvent le style clair organisent également un concours annuel, où l'organisation gagnante se voit décerner un Plain English Award selon différentes catégories liées au genre discursif (brochure, lettre, site web) ou au domaine (juridique, rédaction technique). Ces sceaux, qui estampillent des textes nous ont guidé dans la constitution de nos corpus de textes en PL destinés au public.

2.3 Caractéristiques linguistiques du *plain language*

Après avoir établi un état des lieux du concept et de ses déclinaison dans le monde, nous passons à la caractérisation proprement linguistique et discursive du style clair.

8. Par exemple l'agence de rédaction technique Avec Des Mots <https://www.avecdesmots.com/prestations-langage-clair/>

2.3.1 Genres discursifs concernés

Une des notions linguistiques centrales sur laquelle se fonde notre étude est celle du genre discursif, mentionnée plus haut et définie comme une "pratique discursive qui, pour des raisons historiques et socioculturelles, a reçu un nom qui peu à peu s'est fixé de sorte que les sujets parlants en ont une conception suffisamment claire pour pouvoir le produire et l'interpréter" (Fløttum et al. 2007). De nombreux genres du discours juridique ont fait l'objet d'études en lien avec les principes du style clair, ou ont été réécrits pour correspondre à ses standards, en particulier :

- des textes normatifs, qui créent du droit (Richard 2006), c'est-à-dire d'une part, les textes législatifs et juridictionnels, incluant les lois (*statutes*), traités, les constitutions, conventions, jugement ; d'autre part, les textes juridiquement contraignants, comme les contrats d'assurance et de location, les formulaires et autres, dont la reformulation a été étudiée notamment par Masson et Waldron (1994) ;
- les textes de diffusion des connaissances spécialisées destinés aux non-expert.es, comme les brochures, ou instructions aux membres de jurys populaires étudiées par Charrow et Charrow (1979) ;
- des textes du milieu professionnel juridique tels que les memoranda diffusés en interne, que certains experts utilisent dans leurs enseignements pour encourager les futurs juristes à utiliser les principes du style clair (Hartig et Lu 2014). On peut également inclure dans ces textes les email et lettres échangés dans un cadre professionnel, qui font l'objet de recommandations dans les manuels rédactionnels (Steadman 2013 : 143).

Nous avons ici évoqué les principaux genres qui font l'objet de simplification ou de clarification. Une discussion plus approfondie des genres choisis pour étudier la diffusion auprès d'un public non-expert est présentée plus loin.

2.3.2 Synthèse des recommandations rédactionnelles sur le style clair

Dans cette section, nous présentons une synthèse des recommandations présentes dans les guides rédactionnels pour rédiger en style clair, que ce soit en contexte francophone ou anglophone (bien que davantage de guides soient disponibles en anglais). Elle est le résultat de l'étude de quinze guides rédactionnels de référence visant une langue claire et simple, publiés entre 2001 et 2018. Nous dressons la liste des règles

les plus fréquentes, afin de guider notre choix des phénomènes langagiers à étudier dans la suite de ce travail. Onze des guides rédactionnels étudiés se fondent sur l'anglais, et trois sur le français. Les guides rédactionnels proviennent de différents pays (Etats-Unis, Canada, Australie, France, Royaume-Uni, Nouvelle-Zélande), et ont été publiés par des acteurs institutionnels (Union Européenne, ministères de la justice néo-zélandais, gouvernement britannique, australien) ou associatifs (l'association Plain English Campaign). Ils visent différents publics-cibles : quatre des guides sont destinés au rédacteur.ices de textes législatifs, les autres aux rédacteurs administratifs s'adressant aux usagers non-spécialistes, et différents genres, notamment des textes de lois, lettres administratives, brochures et textes de médiation du discours juridique en ligne. Si la plupart des guides examinés concernent la rédaction législative et administrative, l'un des guides examinés est adressé aux rédacteurs du domaine médical. L'objectif d'une telle diversité est de voir si un ensemble de règles communes à tous ces guides de rédaction *plain language* peut être établi. Le Tableau 2.1 présente donc une synthèse des règles les plus fréquentes dans les quinze guides étudiés. Pour classer les règles, nous nous inspirons des différents niveaux d'analyse utilisés par Gledhill (2011) pour décrire le *Simplified Technical English*.

TABLEAU 2.1 – Synthèse des règles rédactionnelles pour un style clair

Niveau d'analyse	Règle	dans % des guides	Exemple de formulation
Lexique	Remplacer les éléments lexicaux "pompeux" par des alternatives "claires et simples"	100%	Use simple and familiar words. The following is a list of words that can often be replaced by the plainer equivalents shown (Digital Government New Zealand 2020)
Terminologie	Eviter et/ou expliciter les termes spécialisés	100%	Avoid confusing legal and technical jargon (US Department of Transportation 2014)

Acronymes	Expliciter les acronymes	60%	Make sure you define all your acronyms, even familiar ones. (Australian Office of Parliamentary Counsel 2016)
Morphologie	Utiliser des pronoms de 1ère et 2ème personne	27%	Use "you" and other pronouns to speak directly to readers. (US Department of Transportation 2014)
	Préférer les formes verbales aux formes nominales	53%	Les verbes sont beaucoup plus concrets que les noms. (Bureau de la traduction 2018)
	Utiliser le présent	27%	The present tense should be used wherever possible. (Office of the Scottish Parliamentary Counsel 2006)
Groupe nominal	Eviter les cascades de noms	40%	Readability suffers when three words that are ordinarily separate nouns follow in succession. (PLAIN 2011)
	Eviter les prédicats complexes	60%	Many nouns ending in -ion are simply verbs in disguise. Ask yourself if there is a verb that could replace your noun phrase. (Commission Européenne 2011a)

Groupe verbal	Ne pas utiliser le passif ; préférer l'actif	87%	Use the active voice rather than the passive voice. (Australian Office of Parliamentary Counsel 2016)
Proposition Phrase	Éviter la négation	47%	Employer à bon escient la forme négative. (Bureau de la traduction 2018)
	Avoir une structure thème-rhème claire	33%	Longer sentences should not have more than three items of information ; otherwise they get overloaded, and readers lose track. (Plain English Campaign 2001b)
	Limiter la longueur des phrases à 20 ou 25 mots	60%	A good average sentence length ('ASL') is 15 to 20 words. (Plain English Campaign 2022)
Texte	Limiter la taille des paragraphes	13%	Writing experts recommend paragraphs of no more than 150 words (PLAIN 2011)
Mise en page Typographie	Utiliser une mise en page aérée	60%	Designing a simple physical layout makes a document easier to read. (Office of the Scottish Parliamentary Counsel 2006)

	Recommandations typographiques	20%	PUTTING EVERYTHING IN CAPITAL LETTERS IS NOT A GOOD EMPHASIS TECHNIQUE.(PLAIN 2011)
--	--------------------------------	-----	---

Que nous révèle ce tableau sur les caractéristiques centrales du style clair ? Tout d'abord, il suggère que les règles les plus fréquentes fonctionnent en quelque sorte en miroir des caractéristiques centrales du discours juridique présentées au Chapitre 1. Certaines de ces caractéristiques sont transversales aux discours spécialisés (comme la fréquence du passif ou de prédicats complexes), c'est pourquoi le guide rédactionnel lié au discours médical (Plain English Campaign 2001b) contient également des recommandations similaires.

En examinant les recommandations qui apparaissent dans plus de 50% des guides rédactionnels de notre synthèse, on note en premier lieu que les plus fréquentes concernent le lexique, qui doit rester simple et familier. Cette notion de vocabulaire "simple" et "familier" repose sur une certaine conception de la complexité lexicale et n'est pas sans poser problème, comme nous le verrons, mais il s'agit de la recommandation qui apparaît dans tous les guides rédactionnels. La terminologie spécialisée doit également être évitée lorsque le texte s'adresse aux non-spécialistes. Or, nous l'avons vu, à la fois le lexique et la terminologie sont particulièrement spécifiques et nécessaires pour comprendre un texte dans le domaine du droit. Il est donc logique que les guides rédactionnels mentionnent tous des recommandations sur cette question. Les guides publient souvent des mots à éviter et à leur substituer, comme on le voit dans le tableau ; l'influence du développement des langues contrôlées est ici claire, puisque celles-ci se fondent en général sur des règles ayant pour but d'influencer la sélection d'éléments lexicaux, et portant notamment sur l'usage des acronymes (O'Brien 2003). Ils recommandent, d'une part, d'éviter la répétition et d'éliminer systématiquement les mots et termes redondants ; or, nous avons vu que le discours du droit, en particulier législatif, a pour caractéristique discursive d'être particulièrement redondant. D'autre part, les guides rédactionnels préconisent de ne pas utiliser de synonymes, c'est-à-dire deux mots différents pour un même concept, par exemple "*senior citizens*" et "*the aged*" (PLAIN 2011), la synonymie et les répétitions étant présentés comme une source d'ambiguïté (Commission Européenne 2011a,b) et de confusion (PLAIN 2011) pour les lecteurs et lectrices.

Pour rédiger du contenu destiné aux non-spécialistes, les rédacteurs.ices des ins-

titutions juridiques sont également encouragées à utiliser les pronoms de première et deuxième personne (cf. ligne 4 du Tableau 2.1) pour renforcer la relation interpersonnelle entre le texte et son lectorat, pour aller à l'encontre du caractère impersonnel du discours juridique, qui permet dans le discours spécialisé d'atteindre le niveau de généralité et d'abstraction nécessaire à la loi et à la justice. Cette règle rappelle donc la notion de *conversationnalisation* de la communication entre les institutions juridiques et administratives et son lectorat (Turnbull 2018).

Une autre des préconisations fréquentes pour rédiger en style clair (présente dans 60% des guides rédactionnels étudiés) est de limiter le nombre de mots par phrase, afin d'aider les lecteurs.ices à en absorber le contenu : *Shorter sentences are also better for conveying complex information; they break the information up into smaller, easier-to-process units* (US Department of Transportation 2014). Il est donc recommandé de limiter le nombre de mots par phrase, de 15 à 25 mots en moyenne en anglais, de 20 à 25 mots en français. Il semble qu'il y ait un consensus concernant la longueur moyenne des phrases à atteindre :

Most experts would agree that clear writing should have an average sentence length of 15 to 20 words. This does not mean making every sentence the same length (Plain English Campaign 2022).

A nouveau, les guides rédactionnels émettent cette recommandation en opposition aux caractéristiques des discours spécialisés, puisque nous avons vu que les différents genres discursifs juridiques peuvent contenir des phrases de plusieurs dizaines voire centaines de mots (Bhatia 1983). Les guides suivent ici à nouveau l'exemple des langues contrôlées : limiter la longueur des phrases à 20 mots et avoir une longueur moyenne de phrase à 25 mots est également typique de certaines langues simplifiées, comme l'AECMA Simplified English (Farrington 1996).

La plupart des manuels découragent également l'utilisation de ce que nous désignons comme prédicats complexes, composé de verbes suivis de noms prédicatifs (ou noms déverbaux). Certains guides expliquent en détail la formation morphologique de ces noms prédicatifs, souvent reconnaissables par leur suffixe (-ion et autres), utilisés notamment pour réaliser un Procès sous forme nominale (Plain English Campaign 2001b). L'allongement des phrases qui découlerait de l'emploi des prédicats complexes est également pointé du doigt (Commission Européenne 2011a). Plusieurs manuels fournissent donc une liste de prédicats complexes contenant des noms déverbaux (ou noms prédicatifs) à éviter, et leur équivalent non-nominalisé, par exemple *to conduct an investigation* vs. *to investigate* Parliamentary Counsel Office (2021). Le guide en question utilise la métalangue linguistique pour parler

de ce que nous désignons comme des prédicats complexes : il recommande de proscrire la forme verbe "faible" des verbes suivie d'un nom prédicatif, au profit d'un verbe "fort" Parliamentary Counsel Office (2021). A nouveau, on peut noter que plusieurs guides rédactionnels emploient, pour désigner les noms prédicatifs, le terme de *nominalisation*, bien que celui-ci renvoie à un phénomène plus large. D'autres guides rédactionnels n'utilisent pas ces termes, mais parlent de "verbes déguisés" (Commission Européenne 2011b) ou encore de "verbes cachés" (*hidden verbs*) :

Too often, we hide verbs by turning them into nouns, **making them less effective and using more words than we need.** PLAIN (2011)

Les prédicats complexes contenant des noms prédicatifs rendraient "moins efficace" la communication en alourdissant inutilement les phrases - cette préconisation est donc également en partie liée à la question de la longueur des phrases. Les recommandations incluent souvent des exemples de prédicats complexes à remplacer par des verbes dits "forts" : *to make an application, to ensure compliance, to conduct a hearing, to make a discovery* seraient à remplacer par *apply, comply, hear* et *discover*. Elles soulignent également la dimension abstraite des noms prédicatifs, et plus généralement des noms : l'un des guides explique que les verbes sont beaucoup plus concrets que les noms, c'est pourquoi il faudrait remplacer les formes nominales par des formes verbales simples (Bureau de la traduction 2018). Ainsi, dans environ la moitié des guides, les groupes nominaux complexes composés de plusieurs GN, appelés "cascades de noms" ou "noms en cascades" en français, sont également dans le collimateur. En anglais, ils sont appelés "*noun sandwiches*" (US Department of Transportation 2014) ou "*noun strings*", et définis comme une succession de noms dont certains ont une fonction adjectivale ; ils doivent être limités à trois noms maximum en anglais (PLAIN 2011). On voit ici à nouveau un écho des caractéristiques du discours juridique, dont on a vu au Chapitre 1 qu'on évoque souvent son style nominal (Crystal et Davy 1969).

De même, les règles du PL visent à exclure, à moins qu'elle soit vraiment nécessaire, l'utilisation du passif, pour des raisons similaires, comme le montre cet extrait du guide rédactionnel publié par l'association britannique Plain English Campaign :

[Passive sentences] can be confusing. They often make writing more long-winded. They make writing less lively. (Plain English Campaign 2022)

Les raisons avancées pour cette restriction de l'usage du passif sont d'abord liées à la notion de longueur, comme pour le vocabulaire non-approuvé : le passif étant composé de plus de mots qu'une phrase à l'actif (car le groupe verbal au

passif contient non seulement le verbe lexical mais aussi l'auxiliaire), ce dernier est vu comme le support d'énoncés plus clairs et plus concis (Government Digital Service 2016). L'idée qu'une phrase longue est nécessairement plus complexe qu'une phrase courte se retrouve également, nous le verrons, dans les différents scores de lisibilité qui ont été développés, parce qu'un énoncé long demanderait plus d'effort de mémorisation, et donc un surcoût cognitif. Par ailleurs, le passif serait la source de "confusion". C'est en effet une question de structure informationnelle qui motive cette recommandation, les auteurs des instructions indiquant explicitement que la voix active a pour but de placer le justiciable ou l'administration (celui qui "fait") en premier dans la proposition, c'est-à-dire de faire coïncider le rôle sémantique d'Agent avec la position thématique de sujet. Nous revenons en détail sur cette analyse du passif au Chapitre 3.

Enfin, il est notable que les règles pour rédiger en style clair ne portent pas uniquement sur des traits purement linguistiques et textuels, mais également sur des questions de mise en page et de multimodalité. Une place importante est donnée à la mise en page et à la présentation du contenu des textes, qui doivent être claires et espacées, et veiller à la proportion de texte et d'espaces blancs pour faciliter la lecture. La plupart des guides étudiés recommandent également de présenter les informations sous forme de listes et de tableaux, pour expliquer de manière plus synthétique, en faisant apparaître des relations que les lecteurs ne verraient pas sinon (PLAIN 2011). De même, plusieurs règles portent sur la structuration en sections, paragraphes et sous-paragraphes, et également sur l'utilisation de titres et sous-titres. Dans les textes destinés au grand public, les paragraphes doivent rester courts (pas plus de 250 mots), sans être trop courts (PLAIN 2011), afin de montrer explicitement l'organisation du contenu aux lecteurs, ainsi que les idées importantes (US Department of Transportation 2014). Pour faciliter la recherche d'information dans le texte, il est conseillé d'employer des titres pour indiquer le thème d'un paragraphe ou d'une section, afin de permettre aux lecteurs de trouver rapidement l'information dont ils ont besoin. Enfin, la typographie est également source de régulation pour rédiger en style clair : le soulignage est déconseillé et les guides rédactionnels peuvent contenir des recommandations sur la typographie, par exemple l'utilisation de lettres majuscules, comme on le voit dans le dernier exemple du Tableau 2.1, non dénué d'humour. La mise en page et la dimension multimodale sont donc également une constante des guides rédactionnels. Puisque nous nous intéressons uniquement aux caractéristiques linguistiques, cette multimodalité a été exclue de notre étude ; elle est cependant un élément crucial, auquel les défenseur.euses du PLM y consacrent beaucoup d'espace dans les ouvrages de référence (Cutts 2013).

2.3.3 Variabilité des recommandations pour un style clair

Les règles centrales du style clair que nous venons de décrire brièvement semblent, pour certaines, semblables ou proches de certaines règles des langues simplifiées. Cependant, de nombreuses variations apparaissent dans les différents guides rédactionnels.

Variations selon le degré de spécialisation

Il y existe tout d'abord des différences liées au genre discursif et au degré de spécialisation : les recommandations spécifiquement liées à la rédaction de textes législatifs ne sont pas nécessairement pertinentes pour les textes de diffusion aux justiciables. À l'inverse, les recommandations liées par exemple à la taille des paragraphes ou à la longueur des phrases ne sont pas nécessairement appropriées pour les textes de lois, traditionnellement divisés en articles ou sections et sous-sections, d'après l'un des guides : "*In legislation, with the possibility of shredding⁹ to help the reader follow what is being said, the number of words in the whole sentence is not so relevant.*" (Parliamentary Counsel Office 2021). De même, si les guides rédactionnels destinés aux législateurs indiquent que les pronoms doivent être utilisés avec parcimonie (voire si possible ne pas être utilisés), notamment dans le but de rester neutre en termes de genre (*gender-neutral*), les manuels adressés aux rédacteur.ices qui diffusent le droit auprès des justiciables, quant à eux, leur suggèrent d'employer les pronoms de première personne du singulier WE et NOUS pour désigner les rédacteurs eux-mêmes où l'agence pour laquelle ils écrivent, et YOU et VOUS, dans le but d'apporter une "touche personnelle" (Office of the Scottish Parliamentary Counsel 2006). Si ces guides encouragent d'apporter une touche personnelle (grâce aux pronoms) et d'employer un lexique simple, ceux destinés aux législateur.ices découragent à l'inverse l'utilisation d'argot et d'un ton familier, qui, nous l'avons vu, relève d'un registre soutenu, comme le confirme l'un des guides rédactionnels : *The guideline that says to use the simplest word that conveys the meaning does not mean that we may use informal or jargon words in legislation. Our vocabulary is firmly at the formal and conservative end of the continuum.* (Parliamentary Counsel Office 2021). Concernant le registre de langue, l'émphase est donc mise sur le lectorat visé, dont dépend le ton employé. La question de l'utilisation de formes contractées de l'anglais, plus familière, est par exemple évoquée dans certains guides en anglais ; dans ce cas, c'est encore les lecteurs qui doivent guider le choix du rédacteur : "As

9. C'est-à-dire de découper l'information en sections, sous-sections et/ou articles, alinéas.

in any other type of writing, it's important to focus on your reader. If you would speak more formally to someone, then you should probably write to them the same way, without contractions" (PLAIN 2011). Dans les textes de médiation juridique, la question du registre de langue peut être résumée ainsi :

Conversational (if appropriate), but neither colloquial, nor "breezy" (Petelin 2010).

Cette recommandation, liée à la dimension communicative, invite à observer si les rédacteur.ices qui s'adressent aux justiciables en contexte anglophone incluent des marques d'oralité. Or, cette recommandation liée à la dimension communicative de la diffusion des discours spécialisé n'apparaît pas dans les guides en français en tant que telle, ce qui suggère également des divergences selon la langue visée par les recommandations.

Variation selon la langue

En effet, certaines variations sont spécifiques à la langue de rédaction. Certaines recommandations pour simplifier l'anglais ne conviennent pas pour le français, comme le suggère un des guides :

The experience of trying to use both plain English and plain French has had some interesting consequences – for example, experiments with the technique of setting out headings in the form of questions revealed that it is not particularly effective in French where questions tend to be lengthier. (Office of the Scottish Parliamentary Counsel 2006)

De même, cinq guides rédactionnels anglophones conseillent d'utiliser le présent, estimé plus immédiat que le futur et le conditionnel (US Department of Transportation 2014). Ce n'est pas le cas des guides francophones, où seul l'un d'eux conseille d'utiliser les temps appropriés pour "respecter l'ordre chronologique des faits, actions, événements évoqués" (Bureau de la traduction 2018). Cette variabilité est bien sûr prévisible, puisque le style clair doit s'adapter aux spécificités linguistiques de la langue à laquelle il s'applique.

Variations institutionnelles

Enfin, l'un des facteurs de variabilité des règles du style clair provient de variations institutionnelles. Richard (2006) montre par ailleurs que les recommandations

concernant l'emploi de *shall* divergent dans les guides rédactionnels à l'intérieur d'un même pays, les Etats-Unis : si son utilisation pour l'écriture de textes législatifs est encouragée dans les mêmes contextes que celle de *must* dans le Minnesota, on recommande seulement son emploi pour exprimer une obligation dans l'Alaska ou le Delaware, tandis que le guide rédactionnel fédéral (PLAIN 2011) décourage toute utilisation de ce modal, considéré comme archaïque et ambigu¹⁰. On note toutefois que certains guides recommandent de l'utiliser pour exprimer les obligations, et précisent à l'inverse que *shall* et *must* ne sont pas interchangeables (c'est le cas notamment de l'*Alabama Legislative Drafting Style* cité par Richard 2006).

Environ la moitié des guides rédactionnels comportent des règles sur la négation : certains recommandent de supprimer et de ne pas employer de phrases contenant deux (ou trois) formes négatives (*double negatives*), quand d'autres préconisent de privilégier des phrases déclaratives à la syntaxe positive plutôt pas négative, la première étant considérée plus intelligible que la seconde (Office of the Scottish Parliamentary Counsel 2006). Des règles sur la structure thème/rhème viennent restreindre en particulier l'utilisation de phrases complexes, mais seulement dans un tiers des guides. On retrouve également ce type de règles dans un des guides pour le français : "N'enterrez pas d'informations importantes au milieu d'une phrase !" (Commission Européenne 2011b). On notera que, contrairement aux règles rédactionnelles pour les LC, certaines catégories ne sont représentées que dans peu de guides rédactionnels : ainsi seuls quatre guides contiennent des recommandations sur la syntaxe des adverbes et modifieurs, par exemple : "*Keep modifiers close to the words they modify.*" (Australian Office of Parliamentary Counsel 2016). Enfin, un seul guide rédactionnel contient des règles pragmatiques plus pragmatique, déconseillant l'emploi de métaphores et de clichés (Office of the Scottish Parliamentary Counsel 2006).

Plusieurs facteurs peuvent expliquer ces variations institutionnelles. Tout d'abord, les guides rédactionnels pour un style clair varient selon le niveau de détail des règles. Certains faisant plusieurs dizaines de pages et contiennent des règles extrêmement détaillées, reposant parfois sur des descriptions grammaticales très précises ou s'appuyant sur des études de psycholinguistique. D'autres, à l'inverse, ne sont composés que de quelques pages et sont parfois insérés dans des guides stylistiques plus généraux de l'institution à laquelle ils appartiennent. A la suite d'O'Brien (2003), on peut en conclure que les variations dépendent d'à quel point l'ensemble des

10. Décrit assez durement : "*Shall*" is one of those officious and obsolete words that has encumbered legal style writing for many years. The message that "*shall*" sends to the audience is, "*this is deadly material.*" "*Shall*" is also obsolete. (PLAIN 2011)

règles est considéré comme rigide. Les guides rédactionnels du gouvernement fédéral américain (PLAIN 2011) et des conseillers parlementaires écossais, très longs, détaillés et référencés, contiennent par exemple des règles bien plus longues et précises que les quelques pages publiées par le gouvernement néo-zélandais, qui contiennent, pourrait-on dire, les recommandations essentielles pour un style clair. Un second facteur qui peut expliquer est une forme de subjectivité : selon l'institution, certaines caractéristiques peuvent ne pas être perçues comme centrales pour rédiger en style clair. Cette variabilité est à garder à l'esprit dans l'observation des productions réelles des institutions.

Ces variations nous rappellent que le *plain language*, bien que que partiellement inspiré des langues contrôlées, n'est pas standardisé, comme le rappelle Petelin (2010). Cela entraîne des différences dans les règles proposées dans les guides rédactionnels, où peu de recommandations sont finalement partagées par tous les guides rédactionnels, ce qui permet d'établir les caractéristiques centrales qui forment l'essence du PL et que nous avons présentées ci-dessus.

Cette variation confirme que le PL correspond donc non pas à une langue contrôlée standardisée comme le STE, mais davantage à un style rédactionnel :

Plain language [can be defined] as : " a deliberate linguistic style, consciously adopted with the rhetorical intent of making specialized knowledge clearer and more accessible to non-experts" (Gledhill et al. 2019).

Plusieurs auteurs insistent en effet sur ce point : le PL est un ensemble de recommandations, mais non de règles comme nous l'avons vu plus haut chez Petelin (2010) ; c'est d'ailleurs pourquoi nous préférons employer le terme de style clair dans ce travail. Cependant, les recommandations récurrentes nous ont permis de dégager les caractéristiques centrales du style clair.

2.3.4 Exemples de textes estampillés "style clair" avant et après simplification

Dans cette section, nous présentons brièvement quelques extraits de textes juridiques réécrits en style clair afin d'illustrer ces caractéristiques. Certains textes sont issus de publications gouvernementales officielles de lois ou traités réécrits en style clair ; d'autres sont des extraits publiés par des associations ou organisations qui récompensent les entreprises, gouvernements, média ou institutions qui s'efforcent

d'appliquer les principes de simplification de la langue dans la communication avec le grand public.

Le premier extrait est tiré de l'étude de Richard (2006) qui étudie l'impact du PLM sur plusieurs textes de loi, et en particulier le remplacement de *shall*, considéré archaïque, par d'autres modaux ou structures modales après l'instauration de la langue contrôlée. L'auteure présente plusieurs extraits de textes juridiques avant et après simplification au prisme du PL. Elle prend l'exemple des traités internationaux et, comme dans le Tableau 2.2, de deux versions de la constitution sud-africaine, l'une intérimaire datant de 1994, et sa version claire et simple datant de 1997.

TABLEAU 2.2 – Extraits de la constitution d'Afrique du Sud avant et après simplification

Constitution de l'Afrique du Sud, 1994 (avant simplification)	Constitution de l'Afrique du Sud, 1997 (après simplification)
South African citizenship shall, subject to Section 20 read with Section 33 (1), be regulated by an Act of Parliament.	National legislation must provide for the acquisition, loss and restoration of citizenship.

On remarque dans le Tableau 2.2 que la reformulation du texte, conformément aux recommandations pour un style clair, se fait par la substitution de *shall*, jugé complexe et archaïque, par *must*, le "modal privilégié de l'obligation en anglais courant" (Richard (2006)), mais passe aussi par une réorganisation complète de la structure informationnelle de la phrase. La version non simplifiée comporte non seulement une structure passive, à éviter, mais également un circonstant en incise qui renvoie à une autre partie du texte et est inséré entre le modal et le reste du groupe verbal, ce qui rend la lecture non-linéaire. Dans la seconde version, la référence intertextuelle est tout simplement supprimée, et la forme passive devient une structure active, inversant la structure thématique de la phrase. La phrase est également plus courte, passant de 19 à 12 mots. On note cependant que la phrase de droite, simplifiée, contient des nominalisations, en particulier *acquisition* et *restoration*.

Le second texte réécrit en style clair présenté dans le Tableau 2.3 est un court extrait d'une fiche d'information néo-zélandaise ayant remporté le *Plain English Awards* pour la meilleure transformation avant/après d'une phrase. Le jury qui a récompensé ce texte souligne la réduction du nombre de termes spécialisés. On remarque également que le groupe nominal complexe composé de trois noms a été supprimé, ce qui est cette fois-ci davantage en accord avec les recommandations qui encouragent à privilégier un style verbal. Entre autres, le mot d'origine latine "*detect*" devient "*find*" en version claire et simple, et la référence au texte de loi disparaît, de même que les deux formes passives "*may be taken*" et "*being detected*".

TABLEAU 2.3 – Gagnant de la catégorie "Best Plain English Sentence Transformation", Plain English Awards, Nouvelle-Zélande, 2018

Texte original	Texte réécrit
If we detect a problem with your file as part of an information matching programme, we must, under section 103 of the Privacy Act, tell you. We must also tell you of the actions that may be taken as a result of the problem being detected from the programme. The Privacy Act calls these actions "adverse actions".	If we find the information we have about you is different from another agency, we may need to change the services we give you. We'll write to let you know about this before we take any action. If you think the information [Department name] gave us is wrong, please get in touch with us so we can talk about this.

Les exemples ci-dessus semblent donc être conformes, à différents degrés, aux recommandations. L'un des objectifs est de proposer des outils systématiques, dans une étude en corpus, si ces règles sont respectées dans des textes produits par des institutions qui proclament utiliser le style clair.

2.3.5 Des textes simples ou simplifiés ?

Il convient, avant de conclure ce chapitre, de distinguer entre différents types de textes de communication entre expert.es et non-expert.es. Les exemples ci-dessus sont des reformulations de textes destinés au public, c'est-à-dire qu'ils sont la réécriture des textes "primaires" rédigés dans un style juridique. Or, parmi les réécritures en style clair, on peut en distinguer plusieurs types de textes :

- Les textes simplifiés qui font référence explicitement au texte original spécialisé, catégorisés comme une "traduction intralange" (Bhatia 1983), comme par exemple la version en style clair de la Convention Européenne sur les Droits Humains, les textes procéduraux de la DILA ou les résumés des décisions de la Cour Suprême du Canada mentionnés plus haut. Ce sont de véritables simplifications s'appuyant et renvoyant systématiquement à des textes législatifs spécialisés
- Des textes fournissant des explications sur un texte de loi spécialisé et servant d'aide à la lecture pour les non-spécialistes, comme le *NHS Constitution Handbook* de 2015¹¹.

Dans ce dernier cas, la reformulation servirait donc plutôt un objectif de "facilitation" de la lecture, pour traduire le terme d'*easification* utilisé par Bhatia (1983) :

In simplification the focus is on the text, whereas in easification it is on the learner.
(Bhatia 1983)

11. On peut y ajouter un type de texte qui n'est pas dans notre corpus mais digne d'être mentionné, celui des *Explanatory notes* publiées par le parlement britannique pour aider les lecteur.ices à comprendre les textes de lois sans en fait partie, comme le précise Williams (2004)

Par ailleurs, certains textes de dissémination des connaissances juridiques sont directement rédigés en langue simplifiée, sans faire référence explicite et systématique à un texte spécialisé, mais en s'appuyant tout de même sur un discours-source. C'est le cas pour la plupart des textes du Citizens Advice Bureau, l'une des sources majeurs de textes pour notre corpus, qui ne référencent pas systématiquement à un texte spécialisé.

Tous les textes de diffusion du droit n'ont donc pas le même statut par rapport aux textes juridiques, c'est pourquoi, plutôt que de parler de "traduction intralingue" (Bhatia 1983), on parlera davantage de discours de simplification, pour explorer les textes se réclamant de la reformulation du discours du droit, car il s'agit d'un discours de médiation qui s'appuie sur un discours premier, spécialisé, même lorsqu'il n'y a pas de texte juridique de référence. Nous reviendrons sur ces différences entre les textes rédigés directement en langue claire et ceux correspondant à des réécritures lors de la discussion de l'élaboration des corpus de travail.

Il est d'ailleurs intéressant de noter qu'une partie du débat parmi les spécialistes concernant le *plain language* concernait les versions simplifiées des textes et les limites des stratégies de reformulation. Les conclusions de certains articles portant sur la simplification de textes juridiques comme les contrats grâce aux principes du PL (Masson et Waldron 1994) posent la question du degré d'accès aux connaissances spécialisées, et se demandent si l'exposition à un seul texte, même réécrit en style clair, permet vraiment de comprendre le système de connaissances et les concepts juridiques. L'une des idées mises en avant par les juristes contre le style clair est que les concepts juridiques sont trop inaccessibles, trop sujets à interprétation, pour être paraphrasés et définis en isolation dans un texte. Un style clair suffit-il à faire comprendre le domaine spécialisé et ses concepts lorsqu'on n'a jamais suivi d'enseignement ou lorsqu'on n'a pas eu l'expérience de ce domaine, si le système conceptuel spécialisé n'est que partiellement expliqué dans le document (Masson et Waldron 1994) ? Un de nos objectifs est d'observer en corpus comment ces difficultés et ces questions sont traitées par les institutions qui se réclament du style clair.

2.4 Résumé du Chapitre 2

Dans ce chapitre, nous avons défini le style clair comme un standard communicationnel, présenté sous forme de règles rédactionnelles pour aider les juristes et autres rédacteur.ices techniques s'adressant au grand public à diffuser des connaissances spécialisées de manière claire et simple. Nous avons décrit les caractéristiques du PL selon les recommandations officielles publiées par les institutions qui s'en réclament, en français et en anglais. Il convient maintenant d'interroger les fondements conceptuels et théoriques de la complexité et de la clarté qui se cachent derrière les règles du style clair décrites ici. Quelles influences théoriques peut-on voir dans la façon dont sont formulées ces règles ? Quelles notions de la lisibilité et de la clarté en linguistique ont façonné le PL ? C'est à ces questions que nous allons maintenant nous efforcer de répondre.

Chapitre 3

Complexité et lisibilité en linguistique : définitions, approches et mesures

Après avoir caractérisé le discours juridique et sa diffusion, il faut maintenant définir ce que recouvrent les termes "complexité" et "clarté" pour les organismes qui se réclament du style clair. Dans ce chapitre, nous nous attacherons à répondre aux questions suivantes : Quelles conceptions de la complexité et de la clarté existent dans la littérature ? Quelle vision de la complexité linguistique et discursive retrouve-t-on dans les guides rédactionnels publiés par les institutions et organismes qui prônent l'emploi du style clair ? Comment cette notion peut-elle être outillée ? Comme nous allons le voir, la complexité linguistique est un phénomène protéiforme, qu'il est difficile de réduire à une seule définition ou à un seul aspect.

Ce chapitre est organisé comme suit : nous présentons un panorama des définitions et approches de la complexité en linguistique, qui sert ensuite de point de départ pour discuter d'abord des différentes perspectives sur la complexité adoptées dans les guides rédactionnels, mais aussi plus généralement des autres notions et outils linguistiques qui ont été développés pour mesurer la complexité, en particulier dans le domaine de l'acquisition et du automatique des langues (TAL). Nous concluons en résumant les concepts et outils sélectionnés dans ce travail pour évaluer le passage d'un corpus de textes spécialisés, considérés complexes, à un corpus de textes proclamés clairs et simples.

3.1 Qu'est-ce que la complexité en linguistique ?

La préconisation d'utiliser le style clair par diverses institutions vise, nous l'avons vu, à donner accès aux connaissances juridiques et à diffuser un discours jugé trop complexe. Pour les défenseurs du style clair, la clarté, qui est traditionnellement opposée à la complexité, est définie de manière pragmatique : un texte est clair si les lectrices peuvent trouver l'information dont iels ont besoin dans le document. D'après nous, la notion de clarté est donc utilisée comme un synonyme de lisibilité, en particulier dans le cas de textes écrits qui nous occupe. Pour accomplir une démarche ou connaître ses droits, les justiciables doivent en effet pouvoir comprendre *immédiatement* les textes produits par les institutions :

Plain language [is] defined as communication that is comprehended the first time it is encountered, and that relies on textual features such as active voice and common terms (Rossetti et al. 2020).

Le style clair repose, comme nous l'avons vu, sur certains traits linguistiques et discursifs qui visent à garantir la lisibilité et la compréhensibilité des textes. Il convient à présent de se demander, à l'inverse, comment sont conceptualisés et théorisés les éléments qui nuisent à une communication efficace et augmentent la complexité d'un texte. Nous interrogeons en particulier l'influence des différents courants linguistiques sur la conceptualisation de la complexité. En linguistique, comment sont conceptualisés et théorisés ces éléments qui sont considérés comme empêchant une communication efficace et augmentent la complexité d'un texte ? Comment les différents courants de la linguistique influencent-ils les mesures de la complexité ?

3.1.1 La complexité, un concept aux multiples facettes

La complexité est un concept qui n'a pas de définition universellement acceptée, puisqu'elle dépend fortement de l'approche employée, ainsi que de la branche de la linguistique qui s'en est emparé. Tout comme le style clair, la complexité reste donc une notion difficile à définir et qui renferme de multiples dimensions. La clarté repose quant à elle sur des éléments textuels préférés et d'autres proscrits (ou en tout cas découragés), que nous avons vus au Chapitre 2 et que nous rappelons ici :

- "interdiction" du passif et préférence pour l'actif ;
- phrases courtes, limitées à 20 ou 25 mots ;

— mise en page efficace.

Pourquoi, linguistiquement, une communication contenant ces traits textuels serait-elle efficace ? Cette question évoque différentes théories de la communication et de la complexité linguistique, notamment (Grice 1975), qui a développé la théorie du principe de coopération sous la forme de maximes conversationnelles. (Foudon 2008 : 63) propose pour une discussion détaillée de ces maximes et du principe de coopération. Les maximes offrent un point de départ intéressant à une discussion de la complexité, puisqu'elles convoquent des problématiques bien connues en lisibilité et dans les langues planifiées. Ces maximes sont classées en quatre catégories : les maximes de quantité, qui stipulent que la contribution textuelle ou conversationnelle doit être aussi informative que nécessaire, mais pas plus ; les maximes de qualité, qui concernent la vérité du propos (dire ce qui est vrai ; ne pas dire ce qui ne peut être établi comme vrai selon des raisons suffisantes) ; la maxime de relation, qui a trait à la pertinence ; et enfin les maximes de manières :

— Évitez de vous exprimer de manière obscure.

— Évitez l'ambiguïté.

— Soyez bref.

— Soyez ordonné.

Les recommandations sur la mise en page et la structuration des textes en paragraphes peuvent par exemple être reliée à la maxime "Soyez ordonné.e". Les maximes de qualité font quant à elles directement écho aux traits et à la visée principale du PL, le style clair s'opposant à une "manière obscure" de s'exprimer. Ainsi la limitation du nombre de mots par phrase renvoie-t-elle à la maxime de concision. La nécessité de la concision est due au coût cognitif entraîné par un texte long et/ou contenant des informations superflue, qui nécessite que les lecteur.ices fournissent un plus grand effort de mémorisation. C'est en partie selon ce même principe que sont découragées les nominalisations et le passif qui rallongent, selon les guides rédactionnels, les énoncés ; ils sont également vus comme augmentant l'"obscurité" et possiblement l'ambiguïté (ce dont nous discuterons plus loin). C'est pourquoi, nous l'avons vu, les LC et le PLM préconisent des textes et phrases courtes, et c'est également pour cette raison la longueur des phrases dans le discours juridique est souvent fustigée.

Un texte ou un énoncé se doit donc d'être court pour être compris et retenu. Cependant, il faut également prendre en compte les maximes de quantité, puisque la question de l'information nécessaire et *suffisante* dans est texte est au cœur des

préoccupations du PLM et de la reformulation du discours juridique pour les justiciables. Les maximes de quantité renvoient au risque d'incomplétude évoqué par Warnier (2018), qui discute les problèmes potentiels dans la rédaction des exigences spatiales : l'omission de certains éléments, par exemple l'argument d'un verbe, peut entraîner un manque d'information et donc une erreur de compréhension d'un texte (Warnier 2018 : 11). L'auteur évoque également un autre risque communicationnel mentionné par Grice, l'ambiguïté, qui peut par exemple être due à l'ambiguïté morphologique, certains mots pouvant appartenir à plusieurs catégories grammaticales (par exemple *state* en anglais, qui peut être nom ou verbe). L'ambiguïté peut également être lexicale ou terminologique, et est alors souvent imputable à la polysémie de certains mots ou termes. Comme nous l'avons vu dans le Chapitre 1, le discours juridique contient de nombreux termes qui sont polysémiques. On peut ainsi prendre l'exemple en droit canadien anglophone du terme *information*, qui dans son sens général est équivalent au français "information" et est non-comptable ; dans ses emplois juridiques en droit criminel canadien, il s'agit d'un terme, qui désigne un document rédigé et signé par un.e agent.e de police sous serment au début de l'instruction, et qui permet de lancer les poursuites contre une personne mise en examen (Coughlan 2020). Le terme, dans ces contextes spécialisés, est un nom comptable, comme on le voit dans les exemples ci-dessous :

- (21) Any one who, on reasonable grounds, believes that a person has committed an indictable offence may lay **an information** in writing and under oath before a justice, and the justice shall receive the information (...). (Code Criminel)
- (22) What is different (canadien) here than in the great run of cases in which informations are withdrawn by the Crown is that the information in this case is that of a private informant, not a law enforcement informant (...) (Cour d'Appel d'Ontario 2010)¹

Ces extraits pourraient être source de confusion pour une personne qui les lirait sans connaître la signification et les emplois grammaticaux spécifiques de ce terme dans le système judiciaire canadien. La difficulté potentielle d'une phrase serait donc imputable à un manque d'explicitation de ce terme polysémique ; en effet, dans le contexte de production spécialisé où il est employé par les spécialistes, il est utilisé par des locuteur.ices qui partagent les connaissances préalables nécessaires pour le comprendre, et pour qui il n'est pas nécessaire de les expliciter, ce qui peut poser des problèmes de compréhension au lectorat non-spécialiste (qui est nombreux, nous l'avons vu, notamment pour les textes législatifs). Les textes produits se heurtent

1. R. v. McHale, <https://www.canlii.org/en/on/onca/doc/>, consulté le 12/05/2022.

donc à deux aspirations contradictoires : la concision et l'explicitation. C'est cette tension qui engendre deux types de complexité différents, selon des linguistes comme Bisang (2014) ou Hansen-Schirra et al. (2020) : la complexité ouverte et la complexité cachée.

Linguistic economy is based on the omission of optional or inherent information, which must be inferred from the contexts, resulting in processing costs, i.e. **hidden complexity**. Linguistic explicitness, however, manifests itself in an increased **overt complexity** recognisable in linguistic structures. (Hansen-Schirra et al. (2020) : 205)

La complexité ouverte, liée au besoin d'être explicite, se réalise par l'ajout de structures linguistiques, tandis que la complexité cachée provient d'une volonté de concision et de brièveté (en conformité avec l'une des maximes de Grice). Pour Hansen-Schirra et al. (2020), le PL se trouve sur un continuum entre les langues de spécialité, qui présenteraient le plus haut degré de complexité cachée (c'est-à-dire d'information implicite) et le plus bas degré de complexité ouverte, et le *Easy language*, équivalent du langage FALC, destiné aux personnes présentant des déficiences intellectuelles ou difficultés de lecture ainsi qu'aux personnes allophones, qui à l'inverse présenterait le plus bas degré de complexité cachée, et le plus haut degré de complexité ouverte. Le PL correspondrait quant à lui une complexité ouverte moyenne. On pourrait arguer que le langage du droit ne correspond pas tout à fait à cette vision des langues de spécialité. Il présente bien, comme nous l'avons vu, un haut degré de complexité caché, avec des termes polysémiques (comme "justice"), des structures et des connaissances implicites partagées par ses usagers. Mais, en anglais au moins, il semble que le langage juridique soit caractérisé aussi par une grande complexité ouverte, puisque, pour le discours législatif, le besoin d'explicitement le périmètre d'application de la loi est crucial, et dans le discours judiciaire, l'explicitation des arguments menant à la décision est également central ; ces deux fonctions ont des conséquences sur les structures linguistiques employées, et donc sur la complexité dite ouverte, comme nous l'avons vu plus haut. Il semblerait donc que le discours du droit présente à la fois une grande complexité ouverte et une grande complexité fermée.

3.1.2 Lisibilité, complexité absolue et complexité relative

Les études sur la lisibilité et la complexité textuelle distinguent également deux types de complexité. Tout d'abord, la complexité absolue :

Absolute complexity is based on an abstract systematic level concerning phenomena like the number of grammatical categories, the fine-grainedness of their distinctions and grammatical relations. (Hansen-Schirra et al. 2020)

Ainsi, comme nous l'avons déjà évoqué, la complexité (et son pendant, la clarté lisibilité) est souvent mesurée en se fondant sur la proportion de certains éléments et structures grammaticales et lexicales. Cette vision de la complexité est au fondement des scores de lisibilité et métriques de complexité lexicales et syntaxiques que nous décrivons plus loin, et a été particulièrement fructueuse dans le contexte du Traitement Automatique des Langues (TAL). François et Fairon (2013) définissent par exemple la lisibilité comme la capacité à "associer un texte à un lecteur" ou une lectrice, grâce à des techniques reproductibles, sans qu'un intervenant humain n'ait à juger la difficulté du texte.

L'autre type de complexité adoptée dans des approches cognitivistes ou psycholinguistiques est la complexité **relative**. O'Brien (2010) explique par exemple que la lisibilité peut être considéré comme une propriété faite de l'interaction entre ce texte et les capacités du lecteur ou de la lectrice à traiter le texte. Dans le cadre d'études psycholinguistiques, elle est mesurée par des outils telles que le temps de lecture capturé par le mouvements des yeux (*eye movement tracking*) (O'Brien 2010)) ou des tâches de compréhension (Masson et Waldron 1994).

Relative complexity pertains to cognitive costs and potentially varies between different individuals. What may be easy for one informant with her/his cognitive properties may be costly for another with different properties. (Hansen-Schirra et al. 2020)

La lisibilité est donc, dans ces approches, résolument *reader-centered*, et dépendante du lectorat qui se retrouve face au texte. Cela nous rappelle que les lecteurs et lectrices du discours du droit ne considèrent pas la clarté de la même manière. Pour un spécialiste du domaine, la clarté, ou la lisibilité, ne sera pas la même chose que pour les justiciables non spécialistes. Pour un juriste, ce sont certaines caractéristiques linguistiques perçues comme complexes par les non-spécialistes, et présentées au Chapitre 1, telles que les répétitions, les expressions binomiales et l'accumulation de GN, qui sont signe de clarté et ont pour but de réduire les ambiguïtés dans l'interprétation du texte par les juges et autres spécialistes, notamment en précisant le sens des notions juridiques employées (Taylor 2003). D'où les résistances des membres de la profession, qui craignent qu'appliquer le style clair aux textes législatifs et aux contrats, notamment, aillent à l'encontre du besoin de précision et de réduction des ambiguïtés du droit. Surtout, cette notion de complexité relative pose la question

du lectorat visé par les textes de diffusion du droit rédigés en style clair. Si cette reformulation du discours juridique s'adresse à tous les justiciables confondus (sauf les personnes dont on a vu qu'elles sont destinataires du langage FALC), quel est le standard d'alphabétisation visé ? Ce groupe, nous l'avons vu, n'est pas uniforme, et contient à la fois des personnes connaissant un vocabulaire étendu (académique) et ayant l'habitude de lire de par leur milieu ou profession, et d'autres ayant un degré moindre d'alphabétisation. Nous verrons dans la suite de ce chapitre comment les outils et mesures de la complexité ont construit un lectorat idéal en ce fondant sur des études de lisibilité et en acquisition du langage. La dimension relative de la complexité linguistique tient non seulement aux capacités cognitives et de lecture du lectorat, mais également au mode de discours, oral ou écrit :

Speech flows along in a temporal sequence; spoken discourse is like a moving current – or, to vary the metaphor, its mode is choreographic, like the movement of a dance. Writing exists in linearized space; written discourse is like a river of ice, frozen in time – its mode of complexity is crystalline and highly compacted. In their relation to the construal of experience, each one makes reality look like itself. (Halliday et Webster 2009 : 77)

Cette métaphore de l'écrit fait écho à la caractérisation typique du discours juridique, dont la dimension formelle et conservatrice rappelle cette métaphore d'une langue comme prise dans la glace, cependant que l'image de la rivière évoque le caractère mouvant du discours juridique. Surtout, cette différence entre les deux modes de discours a des conséquences sur le type de complexité et sa mesure selon le continuum lexique-grammaire. En effet, dans certaines approches comme la GSF, pour faire la construction langagière de l'expérience humaine, il est nécessaire d'une part de nommer les éléments et les choses qui participent de cette expérience, grâce au lexique ; et, d'autre part, d'assigner à ces phénomènes, et aux relations qui apparaissent entre eux, un éventail de propriétés générales, grâce à la grammaire (Halliday et Matthiessen 2013). La complexité est donc habituellement évaluée par des mesures portant sur la grammaire ou le lexique. Cette différence entre complexité syntaxique et complexité lexicale est selon Halliday visible dans les deux modes de discours :

Typically, written language becomes complex by being lexically dense : it packs a large number of lexical items into each clause; whereas spoken language becomes complex by being grammatically intricate : it builds up elaborate clause complexes out of parataxis and hypotaxis.

Halliday distingue donc deux types de complexité : la densité, caractérisée par une plus grande densité d'éléments lexicaux, et l'*intricacy*, qui serait l'interaction

complexe de plusieurs choix syntaxiques systémiques, comme avoir dans la même proposition un passif, une structure impersonnelle et un prédicat complexe. Toujours selon Halliday, les variétés de l'anglais les plus techniques et les plus bureaucratiques, donc celles qui nous intéressent, sont marquées par une grande densité lexicale (Halliday et Webster 2009 : 75). Dans le cas du discours spécialisé juridique, la complexité est *à la fois* lexicale et syntaxique, et relève à la fois de la densité et de l'*intricacy*, selon les caractéristiques que comme nous avons vues aux Chapitre 1. Cette question de la différence dans les modes de discours nous guidera cependant dans l'exploration du passage d'un discours très soutenu à un style plus "conversationnel".

3.1.3 Certains traits linguistiques sont-ils intrinsèquement plus complexes que d'autres ?

Nous avons donc vu qu'il n'y a pas un mais des types de complexité en linguistique. Cependant, nous avons également vu que les guides rédactionnels, les règles des langues contrôlées, et les guides stylistiques généraux contiennent des recommandations communes, liées à des traits linguistiques dont on voit qu'ils sont similaires à certains choix à éviter selon le fameux essai d'Orwell (1946), *Politics and the English Language* :

Never use a long word where a short one will do.
If it is possible to cut a word out, always cut it out.
Never use the passive where you can use the active. (Orwell 1946)

Le fait que ces recommandations soient récurrentes dans diverses recommandations rédactionnelles signifie-t-il qu'il existe des traits intrinsèquement complexes, comme le passif ou les mots dépassant une certaine longueur ? Il semble au premier abord évident qu'un mot plus long ou qu'une phrase avec des mots superflus peut entraîner un sur-coût cognitif et plus grand effort de mémorisation pour le ou la lectrice. Mais la notion de complexité relative que nous avons vue semble contredire cette idée, puis qu'elle est fortement dépendante du contexte de production et de réception. Halliday résume cette idée :

Formulations such as "passive is more complex than active", or "longer sentences are more difficult to process than short ones", are without any value, and not to be taken seriously – it is easy to find contexts where the opposite is the case. A "difficult" text is one that is complex in the wrong way, unrelated to what the situation demands ; or, perhaps, addressed to the wrong audience – such as the wrong age group. (Halliday et Webster (2009) : 77)

Halliday suggère donc que certaines éléments, qui font partie reposent en fait sur des préconceptions de la complexité qui ne seraient pas recevables. Ce sont des considérations liées au contexte de production et de réception des textes, comme le lectorat visé ou l'adéquation des à la situation de communication, qui expliquent la difficulté ou au contraire la lisibilité d'un texte.

La notion de traits lexicaux ou grammaticaux intrinsèquement plus complexes est également influencée par l'approche théorique adoptée. Les recommandations liées aux nominalisations et au passif en sont un bon exemple ; or, comme nous le verrons, toutes les théories linguistiques, et même toutes les langues contrôlées, ne les considèrent pas comme plus complexes. Le PLM véhiculerait donc certaines idées sur la complexité qui peuvent être considérées comme des mythes et/ou stéréotypes sur la complexité. C'est le sens de l'article de (Turfler 2015), qui s'attache à mettre en lumière les enjeux idéologiques qui sous-tendent le PLM (dans le cadre du langage juridique), et montre que les défenseurs du PLM se réclament d'une supériorité linguistique et morale sur le langage juridique traditionnel, considéré comme un bloc linguistique homogène et obsolète.

Plain style is neither more inclusive, nor more consistently effective, nor more democratic than other styles. (...) All styles can be used effectively in legal discourse. (Turfler 2015)

L'objectif de ce chapitre, et plus largement de travail, est d'interroger la conception de la complexité qui sous-tend les recommandations pour un style clair. Par exemple, selon différentes approches, les notions de lisibilité et de complexité ont-elles été opérationnalisées linguistiquement ? Quels outils ont été mis en place pour vérifier que la complexité d'un texte reste en-dessous d'un certain niveau ? En effet, plusieurs règles du style clair, et des études sur la lisibilité, se fondent sur un désir de quantifier la lisibilité grâce à des mesures applicables à n'importe quel texte, qui entraîné la théorisation et la création de mesures de complexité "absolues", pour les différentes strates linguistiques, comme le lexique ou la syntaxe, que nous nous attachons à décrire dans la suite de ce chapitre.

3.2 Approches et mesures de la complexité lexicale

Puisque l'un des traits centraux du style clair est la présence de restrictions portant sur le lexique et la terminologie, nous nous penchons en premier lieu sur la notion de complexité lexicale. Dans les études sur la lisibilité, elle est généralement

considérée sous plusieurs angles. La difficulté du vocabulaire peut être liée à la présence ou absence de listes de mots considérés comme difficiles à comprendre ou peu fréquents, ou liée à la longueur des mots - comme pour les phrases, l'idée est que plus un mot est long (en nombre de syllabes), plus il est difficile à comprendre (Bailin et Grafstein 2001). En termes quantitatifs, pour évaluer une complexité "absolue", on trouve deux mesures principales. La première est la densité lexicale, c'est-à-dire la proportion des items lexicaux (noms, verbes, adjectives et certains adverbes) d'un texte par rapport au nombre total de mots d'un texte, mots grammaticaux inclus, ou par rapport à une autre structure. Pour Halliday & Webster par exemple (2009 : 75), la densité lexicale est le nombre de mots (*tokens*) divisée par le nombre de propositions non-enchâssées, tandis que pour Ure (1971), il s'agit de la proportion d'items lexicaux par rapport au nombre de mots grammaticaux. L'autre pendant de la complexité lexicale est la diversité lexicale. Ce terme est parfois utilisé de manière interchangeable avec le terme de richesse lexicale (*lexical richness*) bien que pour certains auteurs, lexicale n'est en fait qu'un paramètre de la richesse lexicale, qui est multidimensionnelle (Lu 2012). La diversité lexicale est la mesure du nombre de mots lexicaux différents (*types*) utilisés dans un texte. Plus le vocabulaire d'un texte est varié, c'est-à-dire moins le locuteur.ice répète les mots déjà employés, plus sa diversité lexicale est haute (Johansson 2008). Ces deux mesures correspondent à des mesures classiques, dont différentes variantes ont été déclinées dans la recherche sur l'acquisition du langage, dont nous reparlerons plus loin dans ce chapitre. En effet, plusieurs scores de complexité lexicale ont été automatisés et sont calculés principalement selon ces deux paramètres. Nous allons maintenant voir comment cette notion de complexité du vocabulaire est conceptualisée en lien avec le PL.

3.2.1 Vocabulaire approuvé et non-approuvé

L'une des recommandations communes aux quinze guides rédactionnels étudiés au Chapitre 2, en anglais comme en français, est d'utiliser un lexique simple, "familier" et d'éviter les termes ou expressions jugés complexes ou archaïques. On peut donc dire que ces recommandations visent à limiter la densité de certains éléments lexicaux jugés complexes. Surtout, l'influence de LC, le STE et l'ASCMA est flagrante car les guides contiennent tous, sauf un, des exemples et souvent même des listes de mots jugés complexes qui sont à proscrire et à remplacer par leur version dite claire et simple. Les langues contrôlées cités plus haut se fondent en effet sur un lexique restreint, par exemple 156 adjectifs, 97 adverbes, 138 noms et 164 verbes pour l'AECMA (Peterson 1990) ou 850 mots de base pour le Basic English (Ogden

et Halász 1935). Les guides rédactionnels pour rédiger en style clair s'en inspirent donc en décourageant l'utilisation de certains mots et expressions. On distinguera deux types de mots à éviter, résumés dans ce guide rédactionnel francophone :

Pensez à remplacer les mots / expressions qui ne font pas partie du langage courant (**soit parce que ce sont des termes techniques, soit parce qu'ils appartiennent à un parler distingué, élevé, parfois désuet**) par des termes plus simples. (Collette et al. 2002)

Dans cette section nous nous concentrons sur les seconds, les mots et expressions appartenant à un "parler distingué, élevé", c'est-à-dire à un registre soutenu ; la question des termes est discutée plus bas. Les éléments lexicaux que nous appelons "non approuvés" sont désignés par une terminologie différente selon les guides : certains manuels parlent de mots ambigus, d'autres de mots "problématiques" (Australian Office of Parliamentary Counsel 2016) ou encore "pompeux" (*fancy*) (Digital Government New Zealand 2020). Il s'agit à chaque fois de mots ou expressions jugées longs et complexes, appartenant au registre soutenu, et sont à remplacer par des mots jugés simples, familiers, généralement plus courts. La Figure 3.1 présente un extrait de la liste de mots non-approuvés publiée par le gouvernement néo-zélandais, ainsi que leurs substituts clairs et simples. Le tableau suggère que ce sont principale-

FIGURE 3.1 – Mots non-approuvés et leur alternative en anglais

Fancy	Plain	Fancy	Plain
appoint	decide/fix	assist	help
attempt	try	balance	rest
confer	give	data	information
deceased	dead	effected	made/done
endeavour	try	expiration	end
facilitate	help	furnish	give/state/show
grant	give	initiate	begin/start

ment de mots polysyllabiques à l'étymologie latine (par exemple *facilitate*, *deceased*, *attempt*, *endeavour*), dont la version claire et simple proposée est généralement monosyllabique et d'étymologie anglo-saxonne (respectivement *ease*, *dead* et *try*). Ces éléments lexicaux soutenus sont souvent employés dans des contextes plus abstraits que leur alternative claire et simple (par exemple académique). Pourquoi ces mots

sont-ils à éviter ? Il s'agit ici d'éléments lexicaux qui appartiennent à un registre soutenu, c'est-à-dire qu'ils sont moins fréquents que les reformulations proposées, dans les contextes de communication générale, et dans les corpus de langue générale, et qui sont typiquement considérés comme plus complexes, car susceptibles de ne pas être connus par le lectorat non spécialisé, comme le soulignent (Halliday et Webster 2009 : 75). Les études en psycholinguistique ont par ailleurs confirmé que les mots très fréquents sont traités plus rapidement que les mots de basse fréquence (voir Figure 3.1), de même que les mots concrets sont récupérés plus rapidement que les mots abstraits (Hilton 2002).

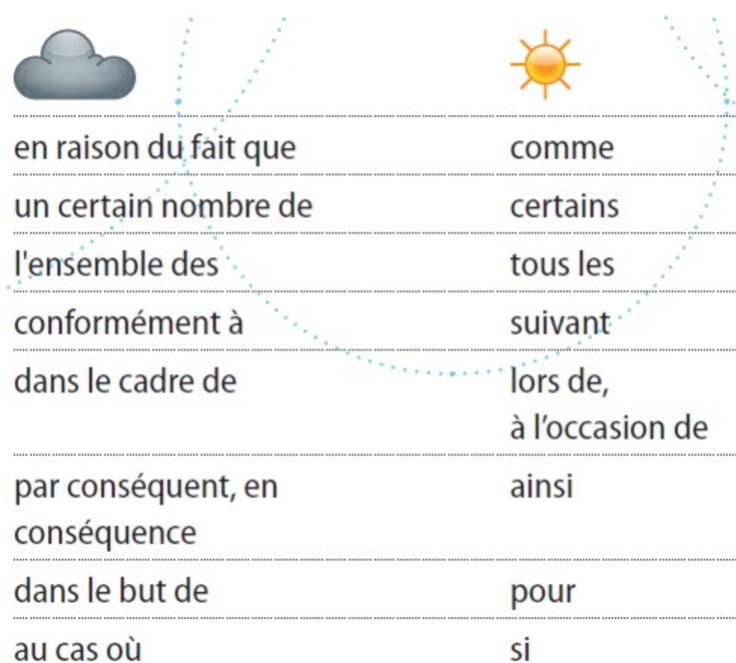
Les modèles d'acquisition du vocabulaire, et donc de complexité ou difficulté lexicale s'appuient en effet, sur la notion de *core vocabularies*, c'est-à-dire la présence ou l'absence dans un texte de mots considérés comme familiers et faciles car très fréquents selon des corpus. Par exemple les 3000 mots les plus fréquents seront considérés comme familiers (Bailin et Grafstein 2001). C'est donc l'influence d'études sur l'acquisition du langage et l'alphabétisation qui est visible dans les guides rédactionnels. La question du *core vocabulary* acquis par un certain groupe d'âge est parfois explicitement mentionnée :

By age 9, you're building up your 'common words' vocabulary. Your primary set is around 5,000 words ; your secondary set is around 10,000 words. You use these words every day. **This is why we tell people to write on GOV.UK for a 9 year old reading age** (Government Digital Service 2016).

Ces recommandations s'appuient notamment sur le concept de profil lexical, ou *lexical profiling*. Les outils de profilage lexical se fondent sur des listes de mots, en particulier les listes des 1000, 2000 mots fréquents que des apprenants d'une langue sont le plus susceptibles de connaître, ou à l'inverse avec lesquels ils sont moins familiers (Laufer et Nation 1995).

On note par ailleurs que les listes de vocabulaire non-approuvé ne sont pas composées uniquement d'items mono-lexicaux. La Figure 3.2 illustre cette idée que le lexique non-approuvé est également constitué d'expressions polylexicales et collocations. Les expressions présentées comme complexes, à gauche, sont notamment des prépositions complexes, caractéristiques du discours juridique, comme nous l'avons vu au Chapitre 2. Ces expressions font partie de la phraséologie du discours juridique. Le guide rédactionnel en question préconise de remplacer ces expressions polylexicales par un seul mot, par exemple de substituer la préposition complexe "dans le but de", la préposition mono-lexicale "pour".

FIGURE 3.2 – Expressions non-approuvées et leur version claire et simple en français (Commission européenne 2011)



en raison du fait que	comme
un certain nombre de	certains
l'ensemble des	tous les
conformément à	suivant
dans le cadre de	lors de, à l'occasion de
par conséquent, en conséquence	ainsi
dans le but de	pour
au cas où	si

S'inspirant des vocabulaires des langues contrôlées et de la recherche en acquisition du langage et en linguistique, les recommandations pour rédiger en style clair suggèrent donc des restrictions au niveau lexical, afin d'éviter ambiguïtés et complexité.

Les mots présentés comme "*plain*" dans le tableau font donc partie d'un vocabulaire qui est considéré comme commun au **lectorat-modèle** ciblé par le site. Les chercheurs en acquisition, de même que les institutions, postulent donc un.e lecteur.ice type avec un certain niveau d'alphabétisation considéré comme "moyen", par exemple à peu près neuf ans d'après la citation ci-dessus afin que le vocabulaire employé dans les textes correspondent au vocabulaire connu par ce lectorat cible. Ce modèle, forcément réducteur, pose cependant problème, puisque les listes de vocabulaire ainsi créées ne prennent pas en compte que la *core vocabulary* varie fortement selon des critères sociaux comme le niveau socio-économique et l'éthnicité (Bailin et Grafstein 2001), mais également le niveau d'alphabétisation, l'âge ou encore la première langue. Concernant en particulier l'application des principes du style clair, Thrush (2001) en étudiant la réception de textes en PL destinés à un lectorat international, a montré que lecteurs francophones et germanophones de textes en *plain English* avaient des difficultés à comprendre les verbes à particules (dont le

sens est non compositionnel) qui avait été substitués aux verbes d'étymologie latine dans le texte original. Le texte devenait plus difficile à comprendre pour des lecteurs dont la L1 est une langue romane romane, contrairement à l'objectif affiché. Malgré tout, bien que ce postulat de départ soit problématique, l'un des objectifs de notre travail est de savoir si ces recommandations lexicales sont respectées, en comparant des listes de mots réputés complexes par les défenseur.euses du style clair, dans nos corpus de textes juridiques vs. notre corpus PLAIN.

3.2.2 La densité terminologique, une mesure du degré de spécialisation ?

Ce n'est pas seulement le lexique soutenu ou peu fréquent qui est visé par les recommandations rédactionnelles, mais aussi les termes spécialisés, et plus précisément l'abus de ces termes spécialisés, qualifié de *legalese* :

The word "legalese" first appeared a hundred years ago to describe what was seen as the unintelligibility and verbosity of the English jargon used by lawyers. (George 2010).

Les guides rédactionnels recommandent des les éviter lorsque c'est possible :

Avoid jargon, legalisms and foreign language if you can use familiar words or expressions instead. (Australian Office of Parliamentary Counsel 2016)

Parmi les éléments à proscrire, on trouve des termes, mais également des termes appartenant à d'autres catégories grammaticales, comme les fameux adverbes archaïques *henceforth*, *hereafter*, *hereby*, *herewith*². La plupart des manuels, comme pour le vocabulaire, contiennent des listes ou exemples de termes ou phrasèmes (unités phraséologiques) à éviter. Office of the Scottish Parliamentary Counsel (2006) conseille par exemple de remplacer la locution latine *ex proprio motu* par l'expression en anglais *of its own accord*, tandis que la Commission européenne (Commission européenne 2011) propose des définitions à utiliser lorsqu'il est nécessaire d'utiliser le "jargon" de l'Union européenne, par exemple *comitology* ou pour les termes qui ont également des usages généraux, comme *cohesion* ("*approach aimed at reducing social and economic disparities within the EU*").

Or, à notre connaissance, peu d'études portant sur le style clair se sont intéressées à cette notion d'un point de vue quantitatif, c'est-à-dire à la proportion de

2. "that give legal writing its musty smell" (Federal Plain Language Guidelines 2011)

termes dans les textes spécialisés vs. simplifiés, bien que l'hypothèse générale soit que ces derniers contiennent moins de termes que les textes destinés aux spécialistes. Ferraresi (2019) explore la notion de **densité terminologique**, dans une étude qui porte sur différents genres spécialisés du domaine de la sécurité alimentaire, mais aussi et surtout des genres discursifs de vulgarisation portant sur les risques alimentaires, et destinés au grand public. À l'aide de listes de termes *ad hoc*, il compare la densité des termes et établit un profil terminologique de ces textes à partir de corpus destinés à différents lectorats (spécialisés ou non-spécialisés), sur le modèle du *lexical profiling* (Laufer et Nation 1995). L'auteur applique donc la notion de *core vocabulary* à la terminologie, en s'intéressant aux *core terms*, c'est-à-dire aux termes-clés du domaine, ainsi qu'aux termes moins fréquents.

La question du contexte des termes présents dans notre corpus de textes destinés au grand public est également intéressante, afin d'observer si des éléments explicatifs entourent ces termes pour les expliciter. Les études portant sur la médiation des discours spécialisés, par exemple scientifique, mais également médical dans le cadre du PL, ont dégagé des tendances concernant les stratégies d'explicitation des termes :

[These strategies correspond] to the breaking down of compact technical terms and their reformulation as more 'user-friendly' formulations with relative clauses and definitions. (Gledhill et al. 2019)

Le transfert et l'explicitation des termes dans le discours de médiation passent par l'utilisation explicite de marqueurs et de reformulation (par exemple "*in other words*"), l'utilisation de comparaisons et d'exemples, et un emploi fréquent de marqueurs de ponctuation telles que les parenthèses ou les guillemets (Jacobi 1985). Certains termes peuvent en effet paraître évidents aux rédacteurs car ils appartiennent au vocabulaire-clé du droit et ont rejoint le langage courant ("double appartenance" évoquée au Chapitre 1). Ferraresi (2019), montre en effet que la densité terminologique, dont il suppose qu'elle est un indicateur mesurable du degré de spécialisation, peut permettre de distinguer les textes spécialisés des textes vulgarisés, puisque ces résultats indiquent que les textes scientifiques sur la sécurité alimentaire contiennent davantage de termes que les textes destinés au grand public. Les différents sous-corpus destinés au lectorat non-expert ne forment cependant pas un bloc monolithique, et certains genres discursifs contiennent davantage de termes que les autres. Il conclut que la terminologie peut être envisagée comme un phénomène gradable, ce qui est également le cas selon lui pour le degré de simplification (*popularization*). Les institutions et guides rédactionnels pour un style clair semblent adopter cette même vision de la terminologie, puisqu'il est recommandé d'éviter si

possible d'employer des termes lorsqu'on s'adresse aux non-spécialistes, donc il faut si possible que les textes aient une plus faible densité terminologique que les textes juridiques destinés principalement aux spécialistes.

Plusieurs études se sont penchées sur la diffusion de la terminologie juridique (Preite 2016) et sur l'impact sur la lisibilité de l'explicitation des termes juridiques (Greene et al. 2012), mais relativement peu se sont intéressées à la densité terminologique comme marqueur quantitatif du degré de spécialisation. Skibitska (2015) interroge cependant l'impact du lectorat visé - spécialiste ou non-spécialiste - dans le cadre de la traduction de l'anglais de spécialisé du tourisme pour différents lectorats - et trouve que la densité de termes diminue de manière drastique dans les textes destinés aux non-expert.es. L'un de nos objectifs est donc de comparer la densité de termes dans nos différents corpus, afin de savoir si celle-ci diminue dans le passage d'un discours spécialisé à un discours destiné à un lectorat profane, conformément aux règles du PL.

3.3 La complexité dans le groupe verbal

Outre l'enjeu crucial de la diminution la complexité lexicale et terminologique, la reformulation du discours du droit passe également par des stratégies de simplification au niveau syntaxique et en particulier dans le groupe verbal. Dans cette section, nous explorons deux traits lexico-grammaticaux particulièrement cités dans les guides rédactionnels : le passif et les prédicats complexes.

3.3.1 Le passif, une source de complexité ?

L'une des recommandations les plus fréquentes pour simplifier rédiger en style clair est d'éviter, autant que faire se peut, d'utiliser le passif. Il s'agit d'une règle classiquement présente dans les langues contrôlées, qu'il s'agisse du Simplified Technical English ou d'autres, par exemple dans le domaine des exigences spatiales (Warnier 2018). Cette recommandation fait également partie des règles édictées dans les guides stylistiques pour la langue générale (Williams et Bizup 2017), qui déconseillent également l'utilisation du passif. D'où vient cette préférence pour la voix active, et par quelle vision de la complexité linguistique est-elle sous-tendue ?

La complexité des constructions passives canoniques : une influence de la grammaire transformationnelle ?

Le passif est considéré comme une source de confusion pour le lectorat, comme nous l'avons vu au Chapitre 2. Son emploi est également découragé la plupart des LC (O'Brien 2010) et dans les guides stylistiques de langue générale (?). Quelles sont les raisons de cette méfiance envers le passif ? Les guides rédactionnels avancent deux raisons : la longueur du passif (avec son auxiliaire), et la confusion qui découlerait de cette structure. Que nous disent les théories linguistiques sur cette forme ?

Selon Warnier (2018), deux raisons font de la voix passive "la mal aimée des langues contrôlées" (92). Les phrases à la voix passive seraient plus complexes que leurs équivalents à la voix active car elles seraient issues d'une transformation de la séquence canonique Sujet-Verbe-Objet, auquel correspondent dans cette interprétation les fonctions sémantiques Agent-Verbe-Patient. L'influence des langues contrôlées, qui pour la plupart préconisent de ne pas employer le passif est à nouveau visible. Apparaît ici non seulement l'influence des langues contrôlées (Gledhill 2011; O'Brien 2010), mais surtout celle de la grammaire générative transformationnelle, pour qui les constructions passives et actives entretiennent une relation expliquée en termes de dérivation. Observons les deux phrases ci-dessous, la première étant tirée d'un jugement de la CSC :

(23) The judgment of the Court was delivered orally by The Chief Justice. (passif)

(24) The Chief Justice delivered the judgment of the Court orally. (actif)

Dans une approche transformationnelle, ces deux exemples sont interprétés comme sémantiquement équivalents et le premier est vu comme la transformation du deuxième. La phrase à l'actif voit l'ordre syntaxique des constituants en adéquation avec ce qui est considéré comme la structure profonde de la phrase, où l'Agent et sujet syntaxique est le juge en Chef (*Chief Justice*) et l'objet syntaxique correspond au Patient, ici le jugement prononcé par la Cour. Au passif, le sujet grammatical correspond à l'objet de la phrase correspondante à l'actif. Hupet et Costermans (1976) résument ainsi l'interprétation de l'opposition passif-actif par la grammaire générative :

A l'actif, l'ordre des relations grammaticales de surface reproduit l'ordre des relations logiques spécifiées en structure profonde; au passif par contre, ces dernières se manifestent en ordre inversé dans la structure de surface. (Hupet et Costermans 1976)

Dans cette approche, donc, une phrase à l'actif, considérée comme une structure non marquée, demanderait un effort cognitif moindre pour être "décodée" qu'une phrase au passif, qui est vue comme une structure marquée. Cette idée que le passif peut donc prêter à confusion se retrouve dans les guides rédactionnels, et c'est pour cette raison que la voix active, qui conserve une relation supposée directe entre le sujet grammatical et l'Agent de l'action, est préférée au passif :

The active voice identifies in a direct way who is doing what. (Digital Government New Zealand 2020)

Cette idée est commune aux recommandations anglophones et francophones :

Pour placer l'utilisateur ou l'administration ou vous-même en tant que sujet-acteur, efforcez-vous d'utiliser la voix active plutôt que la voix passive. (Collette et al. 2002)

La seconde raison de la préférence pour l'actif est liée au caractère optionnel du complément d'Agent en *by* ou *par* dans les structures au passif. En effet, tout passif n'est pas suivi par un groupe prépositionnel. Nous utilisons ici la nomenclature d'Huddleston et Pullum (2005), qui décrivent, tout d'abord, les structures qu'ils appellent passifs longs (*long passives*), qui contiennent un complément d'Agent introduit par *by*. Il est à noter que ces auteurs n'emploient d'ailleurs pas le terme d'Agent pour désigner ces compléments qui fonctionnent comme complément circonstanciels de moyen (*means adjunct* : ils préfèrent utiliser le terme de complément internalisé, afin de conserver une distinction claire entre fonctions syntaxiques et rôles sémantiques (Huddleston & Pullum 2005 : 1428). Les passifs longs contiennent donc un GP, dont nous voyons deux exemples extraits de textes juridiques ci-dessous (en gras) :

- (25) La qualité de travailleur handicapé est reconnue par la commission des droits et de l'autonomie des personnes handicapées mentionnée à l'article L. 241-5 du code de l'action sociale et des familles.
- (26) Mr. Suter was abducted by vigilantes.

Huddleston et Pullum parlent ici de complément internalisé car *by vigilantes* et *par la commission* sont ici internes au groupe verbal, tandis que dans une phrase équivalente à l'actif, le groupe nominal sujet est à l'inverse externalisé. Ils distinguent ces passifs dits longs des passifs courts (*short passives*), qui ne contiennent pas de complément internalisé en *by* ou *par*, comme on le voit ci-dessous.

- (27) L'abus de confiance **est puni** de trois ans d'emprisonnement et de 375 000 euros d'amende. (FR-LEX)

- (28) Under what circumstances, if any, can the presumption of jurisdiction **be rebutted**? (EN-LEX)

Ce sont ces structures en particulier qui expliquent que le passif est présenté comme un trait syntaxique à éviter dans le contexte des langues contrôlées, puisqu'elles risquent alors de mener à une information "incomplète ou inachevée" (Warnier 2018 : 93), violant les maximes de quantité évoquées plus haut, et entraînant un risque d'ambiguïté. Selon ces approches, l'absence du complément d'Agent peut entraîner une ambiguïté ou un manque de clarté, donc une plus grande complexité cachée (selon la typologie établie plus haut), si le ou la lectrice n'a pas les connaissances préalables pour connaître ou déduire du contexte qui est l'Agent. La phrase risque alors de paraître "incomplète ou inachevée car il n'est pas précisé par qui ou par quoi l'action mentionnée doit être faite" (Warnier 2018). Dans certains contextes, l'identification explicite de l'Agent est une information cruciale à ne pas omettre, par exemple dans le contexte de l'aviation. Cependant, certain.es auteur.ices montrent que pour l'anglais quatre passifs sur cinq sont des passifs courts, car l'Agent n'a pas besoin d'être précisé, soit parce qu'il est évident dans le contexte, soit parce qu'il n'a pas besoin d'être mentionné pour que la phrase soit compréhensible. Les guides rédactionnels précisent d'ailleurs que le passif est parfois acceptable, par exemple quand l'Agent est déjà identifié ou n'a pas besoin d'être identifié (Parliamentary Counsel Office 2021).

Les rédacteur.ices techniques et institutions juridiques répètent cependant ce que nous considérons comme une interprétation générative ou transformationnelle du passif, et contribuent à perpétuer sa mauvaise réputation. Pour les défenseurs du style clair, le passif est en effet perçu comme rallongeant inutilement les phrases, et comme source de confusion par sa modification de la structure informationnelle supposée "naturelle" ou linéaire de la phrase canonique.

C'est donc une certaine vision traditionnelle de la complexité grammaticale qui explique cette règle du style clair, où la voix passive est vue comme une transformation de la voix active et où une phrase sans complément d'Agent est vue comme contenant une information manquante, mais il ne s'agit pas de la seule interprétation existante du passif.

Variantes syntaxiques et morphologiques du passif

Avant d'examiner d'autres approches du passif, il convient de signaler que les passifs pointés du doigt dans les guides rédactionnels ne prend en compte que les structures passives prototypiques, qui n'englobent pas à tous les passifs existants. Les constructions passives mentionnées dans les guides rédactionnels sont celles que l'on peut qualifier de canoniques, dans lesquelles sont inclus un auxiliaire exprimé : le verbe *be*, ou *être* en français. Nous désignons ces formes, décrite par certains grammairiens comme périphrastiques, indifféremment par "passif canonique" ou "passif prototypique".

Or, on peut dans un premier lieu distinguer ces passifs prototypiques des passifs sans auxiliaire appelés par *bare passives* (Huddleston et Pullum 2005 : 1430). Pour désigner ces passifs, il existe plusieurs termes : En anglais, on peut les trouver dans des structures caténatives complexes telles que *have + GN + PP*, par exemple ici :

- (29) Have the copies **certified** by a notary public or embassy or other official and keep those copies on hand. (EN-PLAIN)

Une autre forme de passif non-canoniques, plus courante peut-être et également fréquente en français, est un passif sans auxiliaire qui sert de post-modification dans un groupe nominal.

- (30) S'il s'agit d'[un délit **puni** de dix ans d'emprisonnement], la durée de la détention provisoire ne peut excéder un mois, renouvelable une fois.
- (31) [A person **convicted** of the offence with which Mr. Wong was charged] would become inadmissible to Canada, no matter the length of [the sentence **imposed** on him or her].

Le passif, dans tous ces exemples, modifie un nom à l'intérieur d'un groupe nominal et est lui-même modifié par un syntagme propositionnel ou nominal. Une analyse transformationnelle décrirait sans doute ces exemples en termes d'ellipse. On peut en effet gloser ces propositions passives par des propositions relatives contenant un passif canonique, par exemple pour l'exemple (30), *S'il s'agit d'un délit **qui est puni de dix ans d'emprisonnement*** ou pour l'exemple (31) "*A person **who is convicted of the offence...***". Les termes utilisés pour désigner ces passifs "nus" (*bare*), comme passif réduit, passif non-périphrastique ou passif sans auxiliaire, supposent également que le passif est "dérivé" d'une forme sous-jacente, appelée "pleine", ou

encore périphrastique selon les approches. Des auteurs qui n'adoptent pas une approche transformationnelle, tels que (Halliday et Matthiessen 2013 : 182) évitent par exemple de dire qu'une proposition au passif serait "réduite". Faute d'une meilleure traduction, dans la suite de ce travail, nous désignons les *bare passives* par le terme de *passif sans auxiliaire*. Dans tous les cas, ces passifs sans auxiliaire ajoutent, sans en avoir l'air, de la complexité syntaxique en post-modifiant le noyau à l'aide d'un groupe verbal : ils ajoutent de la complexité "ouverte" à l'intérieur du GN qu'ils modifient. Or, il n'est nulle part fait mention de ces passifs sans auxiliaire dans les guides rédactionnels pour un style clair, en anglais comme en français.

D'autres structures et variations morphologiques du passif sont également ignorées par les recommandations. En effet, les auxiliaires *be* et *être* ne sont pas les seuls auxiliaires possibles pour former le passif. En anglais, c'est *get* qui est souvent mentionné comme auxiliaire permettant la construction de structures passives, comme on le voit ici :

(32) If the application **gets rejected**, appealing will take between 6 and 9 months. (EN-PLAIN)

Ces passifs en *get* sont en général évitées dans un style soutenu, c'est pourquoi l'exemple ci-dessus est extrait du corpus destiné au grand public, non du corpus juridique. Surtout, ces passifs en *get* sont restreints à des verbes dénotant des procès dynamiques, comme ici *reject*, mais ne peuvent être utilisés avec des verbes qui dénotent des procès statifs comme *believe* (Huddleston et Pullum 2005 : 1429). D'autres verbes peuvent servir d'auxiliaires en anglais : *have*, comme vu plus haut, mais également *want* (Albrespit 2007). Pour ce qui est du français, Hupet et Costermans (1976) citent également d'autres verbes pouvant servir d'auxiliaires du passif, en particulier les formes pronominales de *trouver* et *voir*.

(33) Le contrat ne crée d'obligations qu'entre les parties. Les tiers ne peuvent ni demander l'exécution du contrat ni **se voir contraints** de l'exécuter, sous réserve des dispositions de la présente section (...).

(34) Si l'un des père et mère décède ou **se trouve privé** de l'exercice de l'autorité parentale, l'autre exerce seul cette autorité.

Contrairement à l'anglais, ces constructions avec d'autres auxiliaires se retrouvent plutôt dans un registre de langue soutenu.

Enfin, le participe passé n'est pas toujours une condition sine qua non pour former le passif. (Huddleston et Pullum 2005 : 1429) parlent de passifs cachés, ou *concealed passives*, pour les passifs qui contiennent non pas un participe passé, mais une forme de gérondif, comme dans l'exemple ci-dessous :

(35) This draft needs **checking** carefully by the editor.

Cette phrase pourrait en effet être paraphrasée par un passif canonique long : "*This draft needs to be carefully checked by the editor.*" Pour les auteurs, la présence d'un GP en *by* contenant un complément internalisé (complément d'Agent) suffit à établir cette forme comme un passif. Albrespit (2007) évoque quant à elle les formes dites moyennes, qui se trouvent en anglais avec des verbes comme *sleep* et *read*.

(36) This bed **sleeps** three.

(37) This book **reads** well.

Dans ces constructions, le Sujet du verbe à l'actif est interprété comme un Patient / Objet affecté, et l'interprétation de la proposition est donc sémantiquement passive : malgré la morphologie exceptionnellement active du verbe, le sujet est bien ici interprété comme le Patient du verbe. Ces emplois intransitifs de verbes habituellement transitifs (directs ou indirects) sont liés à la notion d'ergativité. Halliday et Matthiessen (2013) expliquent que des verbes tels que *cook*, *close*, *open* peuvent être analysés en fonction des notions d'ergatif et non-ergatif plutôt qu'avec les étiquettes traditionnelles "transitif/intransitif" : le procès exprimé par le verbe est-il amené par le participant, comme dans *The lion woke*, ou par une autre entité, extérieure, comme dans *The tourist woke the lion*(340) ? Dans les exemples (36) et (37), les formes moyennes *sleeps* et *reads* contiennent à la fois des traits ergatifs et non ergatifs.

En français, Rouski (2015) évoque deux formes de passif non-prototypiques. Tout d'abord, une forme qui peut se rapprocher des formes moyennes et de la notion d'ergativité vue ci-dessus, où là encore le verbe est morphologiquement actif. Il s'agit du passif pronominal.

(38) L'article 13 **se lit** comme suit : Une partie peut se retirer du régime coopératif en donnant par écrit un préavis d'au moins six mois aux autres parties.

- (39) Le secret professionnel **s'exerce** à l'égard des tiers quels qu'ils soient, y compris les collègues, sauf lorsqu'ils ont eux-mêmes à connaître des informations en cause.

Selon Rouski, ces emplois de formes pronominales reçoivent une valeur passive lorsque plusieurs critères sont réunis : le sujet est inanimé (dans nos exemples, il s'agit bien en effet d'entités inanimées, un article de loi et un concept juridique abstrait), la structure pronominale est utilisée à la troisième personne du singulier, et l'Agent est non exprimé. Pour l'auteur, l'utilisation de ces formes au présent leur confère une valeur déontique et d'obligation, c'est pourquoi elle est particulièrement utilisée dans le discours juridique. Les structures pronominales pourraient donc être glosées par des passifs courts contenant des modaux, "doit être lu comme suit" et "peut/doit être exercé". Une autre variation syntaxique du passif en français, bien qu'elle ne soit pas liée à la morphologie de l'actif, mérite d'être mentionnée ici, puisque, toujours selon Rouski (2015), le discours législatif en est particulièrement friand. Il s'agit d'une variante syntaxique du passif canonique, le passif impersonnel, dont deux exemples sont donnés ci-dessous :

- (40) Une fois qu'**il est établi** qu'un tribunal a compétence, la doctrine du forum non conveniens requiert ce tribunal d'établir s'il devrait exercer cette compétence.
- (41) LE DEMANDEUR DOIT SIGNALER À L'AUTORITÉ QUI A REÇU SA DEMANDE TOUT CHANGEMENT DE RÉSIDENCE (...) EN TRANSMETTANT AUPRÈS DE CETTE AUTORITÉ LE DOCUMENT (...). **IL SERA DÉLIVRÉ** RÉCÉPISSÉ DU DÉPÔT DE CE DOCUMENT.

Il s'agit bien d'un passif canonique, puisque les structures contiennent le verbe *être* et un participe passé, mais le sujet du verbe au passif est instancié non par le Patient, mais par le pronom impersonnel *il*, car la structure syntaxique est une extraposition. Le passif, qui place dans ces emplois canoniques le Patient du procès en position thématique, lui substitue ici le pronom impersonnel, qui marquerait la dimension universelle de la règle de droit, et serait utilisé en particulier pour mettre en avant le résultat du procès, non les participants (Rouski 2015), ce dont nous reparlerons plus bas.

Ces formes non-prototypiques ou atypiques permettent de conclure que le passif est en réalité caractérisé par une grande variété syntaxique et morphologique, dont l'absence dans les guides rédactionnels confirme que c'est l'interprétation dérivationnelle du passif qui est le plus souvent adoptée. L'absence d'Agent exprimé et la longueur supposée des propositions à la voix passive suffisent à justifier que les

réducteur.ices préconisent de l'éviter. Or, d'autres approches du passif existent en linguistique.

Approches non transformationnelles du passif

En premier lieu, on peut se demander si cette réputation de plus grande complexité du passif est confirmée par des études en psycholinguistique. Si les tâches dans certaines expériences indiquent une plus grande facilité de mémorisation des phrases à l'actif, d'autres études suggèrent que celles-ci ne sont pas toujours préférées par les locuteur.ices. Hupet et Costermans (1976) évoquent par exemple une étude ayant montré expérimentalement que des participant.es francophones préfèrent la phrase passive "*La collision fut évitée de justesse par le motocycliste*" à l'active "*Le motocycliste évita la collision de justesse*", ce qui semble lié à une préférence syntaxique, pour conserver l'unité du couple verbe-adverbe dans le premier cas. Hupet et Costermans (1976) citent également des expériences de mémorisations, où les participant.es, devant mémoriser et restituer des phrases, réintroduisent au passif certaines phrases données à l'actif. Ces études expérimentales viennent donc partiellement contredire l'idée que le passif est forcément plus complexe et spontanément évité par les usagers de la langue.

Surtout, d'autres approches théoriques du passif remettent en question cette idée que le passif est intrinsèquement complexe parce qu'il serait une transformation de l'actif. Il existe un débat sur le statut grammatical du passif. Nous l'avons vu, il est la plupart du temps interprété en termes de voix - aussi appelée diathèse -, et conçu par opposition à la voix active (et, dans une moindre mesure, la voix moyenne). Certain.es linguistes remettent cependant en cause cette conception du passif. Beedham, s'appuyant sur une analyse du russe, de l'allemand et de l'anglais, Beedham (1981) adopte une approche originale du passif comme aspect en anglais. Dans son interprétation, *be* est vu comme un auxiliaire d'aspect, à l'instar des auxiliaires *be* et *have* dans les périphrases d'aspect progressive (BE + ING) et perfective (have + EN). Le participe passé du passif est quant à lui vu également sous un angle aspectuel, là encore comme et le participe présent du progressif le participe passé du parfait (*present perfect*). Selon lui, le passif et les groupes verbaux qui apparaissent au passif ne sont donc *pas* dérivés d'une supposée structure à la voix active sous-jacente, car il n'y a pas de paradigme de "voix" selon lui, ni de relation actif-passif, ni même de catégorie "actif" (Beedham 1987). De même, *by* est une préposition ordinaire dont l'un des sens dans le dictionnaire signifie l'agentivité. C'est pourquoi le groupe propositionnel est optionnel, comme les GP le sont typiquement, par exemple *Mary*

is angry (with John). Selon Beedham, le passif exprime un nouvel état, qui résulte d'une action ou d'un événement précédent exprimé par le procès. Pour lui, le passif n'est donc pas sémantiquement équivalent à l'actif, puisqu'il indique que le sujet entre dans un nouvel état :

Passives thus contain simultaneously a semantic element of state and a semantic element of action. Passives and "actives" are therefore not synonymous, since passives contain a statal meaning which is absent in actives. (Beedham 1987)

L'auteur établit une corrélation entre le passif et l'aspect perfectif, *Someone has broken the doll* (état résultant : *her doll is now broken*), qui exprime conjointement une action et un changement d'état du sujet. Pour lui, cependant, le passif est une construction intransitive où le changement d'état entraîné par l'effet du procès sur le sujet. C'est ce changement d'état qui explique qu'on lui attribue le rôle sémantique de Patient, et non le passage au passif d'une phrase à l'actif. Cette interprétation simplifie grandement l'analyse du passif, et apporte un éclairage nouveau sur la distinction entre *short* et *long passives*. Selon (Huddleston et Pullum 2005 : 1428), les passifs courts n'ont en effet pas d'équivalent à l'actif. Dans une approche transformationnelle, il faut donc supposer un agent-sujet sous-jacent comme *someone* non exprimé dans la structure de surface. Dans l'interprétation aspectuelle de Beedham, à l'inverse, il est "parfaitement naturel" (Beedham 1987) que le groupe prépositionnel en *by* ou *par* soit optionnel. Cette approche est particulièrement intéressante, puisqu'elle se démarque nettement d'une conception traditionnelle du passif. Par ailleurs, cette notion de passif comme ayant une valeur d'état ou une valeur résultative n'est pas sans rappeler l'analyse de Rouski (2015), qui s'intéresse à l'utilisation du passif dans le contexte du registre juridique, ici en particulier législatif, en se fondant sur un corpus de textes de l'Union européenne.

Le passif pose **le résultat, l'état des choses envisagé, voulu, attendu**. Dans cette action l'autorité veut se faire discrète, mais sa volonté est entendue comme inconditionnelle. (Rouski 2015)

Pour l'autrice, le passif est l'une des stratégies qui permet de présenter les résultats des actes de prescription de l'autorité juridique, législative ou judiciaire, qui reste cependant invisible car elle n'est pas nommée sous la forme d'un Agent. La fonction rhétorique du passif est donc d'après elle une stratégie discursive parmi d'autres.

Dans les approches fonctionnalistes telles que celle d'Halliday, le passif, bien qu'il soit interprété en termes de voix, n'est pas expliqué en terme de dérivation, mais

également en termes de stratégies d'organisation de la phrase et de la proposition. Cette analyse est intéressante au regard de la fréquence du passif dans les textes juridiques. Le passif ne serait pas complexe dans ces contextes, mais bien une démarche rhétorique visant à topicaliser les concepts et acteurs juridiques. Dans la théorie de la GSF, la fonction du passif est en effet de thématiser un élément du discours. Banks, par exemple, qui étudie l'évolution du discours scientifique en diachronie, voit l'utilisation croissante du passif dans les articles scientifiques à partir du XIX^{ème} siècle, non comme un moyen d'obtenir une plus grande forme d'objectivité, mais comme une stratégie de présentation de l'information :

The passive can then be seen as a resource for placing an element other than the Actor in material process (and its analogue in other types of process) in theme position. Thus, the supposed impersonality produced by not mentioning the agent is the result of the process of thematisation. I do not think that the motivation is to produce impersonal text, and that as a result the thematic structure has non-agentive items in theme position; it is rather that the motivation is to place certain items in theme position, and that in some cases the resource for doing this is the use of the passive. So, the use of the passive is the result of a more **extensive strategy of thematisation, not the other way round**. (Banks 2008)

Le passif sert d'après cette approche à topicaliser un élément du discours, c'est-à-dire à le placer en position thématique, en début de proposition ou de phrase : le reste de la proposition (le rhème), en position focale, fournit un commentaire sur le thème, ou vient ajouter des informations nouvelles. Les approches fonctionnelles de la langue adoptent donc également une autre perspective du passif, qui n'est pas vu comme potentielle source d'incomplétude ou d'ambiguïté, mais bien comme une stratégie linguistique de cohésion discursive à part entière. La citation suivante de Minton nous paraît particulièrement pertinente pour expliquer la fonction du passif, et peut être étendue au français :

The passive voice is an intrinsic part of the English language and used appropriately to maintain the natural flow of writing and presentation of information, it is an essential feature of good writing. Its function is not to conceal or obfuscate, but to maintain stylistic patterns in the presentation of information that have established themselves for very valid reasons in the English language over centuries of usage (Minton 2015, 9).

L'un des objectifs de notre travail est donc d'examiner si les recommandations sur le passif sont suivies dans les textes du corpus PLAIN. Nous souhaitons ainsi explorer la conception du passif dans les textes effectivement produits par les institutions et associations se réclamant du PL. Ces formes sont-elles évitées dans les

différents textes rédigés en style clair ? A l'inverse, les rédacteur.ices utilisent-ils des constructions passives canoniques et sans auxiliaire librement, se démarquant de l'idée que le passif est nécessairement complexe ? Si des constructions passives sont présentes en grande quantité dans le corpus PLAIN, peut-on trouver des régularités dans les éléments qu'elles servent à thématiser ? Nous nous intéresserons donc à la fréquence du passif dans les corpus LEX et PLAIN, afin de voir si ces structures, canoniques mais aussi atypiques, sont présentes dans les deux corpus, et si elles sont caractérisées ou non par la présence de l'Agent. Notre approche sera davantage liée à une approche fonctionnelle et descriptive qu'à une approche transformationnelle. Nous verrons également si des différences quantitatives et qualitatives apparaissent entre les différents genres discursifs représentés.

3.3.2 Prédicats complexes et procès sémantiques

Une seconde dimension de la complexité verbale que nous avons à cœur d'explorer est la notion de procès sémantiques, en lien avec la question des prédicats complexes, c'est-à-dire des constructions contenant un verbe et un nom prédicatif. En effet, les recommandations pour un style clair préconisent, nous l'avons vu au Chapitre 2, un style "verbal" plutôt que "nominal", mais également un style concret plutôt qu'abstrait, ciblant en particulier les prédicats complexes. Pourquoi ces derniers sont-elles considérées comme plus abstraites ?

Selon Halliday, l'enfant acquiert à différents âge des capacités d'abstraction de plus en plus grandes. Il ou elle peut d'abord généraliser en nommant les êtres, animaux, objets et institutions ayant une forme perceptible (de 1 à 4 ans). De 5 à 10 ans, il ou elle peut comprendre des entités abstraites et perceptibles mais aussi des entités abstraites et symboliques. Enfin, à partir de cette période et jusqu'à l'âge adulte, il ou elle développe la capacité de comprendre la métaphore grammaticale, c'est-à-dire quand la construction du sens se fait par chevauchement des catégories grammaticales : les noms ne représentent alors plus uniquement des entités, ni les verbes plus nécessairement des procès. Les nominalisations correspondent donc au plus haut degré d'abstraction puisque les procès y sont représentés sous forme nominale et non sous forme verbale : il s'agit bien d'une métaphore grammaticale. C'est notamment par l'utilisation abondante de ces formes que le discours juridique est considéré comme complexe pour l'utilisateur en termes d'abstraction parce que, à l'instar du discours scientifique, il renouvelle les ressources lexico-grammaticales de la langue (Webster & Halliday 2009).

Nous nous intéressons aux prédicats complexes en lien avec la question des rôles et procès sémantiques. Les procès sémantiques, tels qu'ils sont conçus dans le modèle de la GSF, sont liés à la manière de dont la représentation de l'expérience du monde physique (externe) et psychique (interne) est réalisée par les ressources grammaticales, ici le prédicat. Nous décrivons quatre catégories principales de procès sémantiques, fondées sur les catégories de Halliday & Matthiessen (2013 : 213-126) et la traduction de Caffarel (2009). Tout d'abord, on trouve les procès matériels, qui encodent des sens liés à l'action (*make*) au transfert (*give*) et à la création (*draft*), mais également à ce qui affecte une personne ou une chose (*discriminate against*) ou à ce qui arrive (*occur*). Un procès matériel peut également être abstrait (*treat*), ce qui est fréquent dans le cas du discours juridique, comme nous le verrons. Les participants des procès matériels sont l'Acteur (*Actor*), c'est-à-dire la personne/chose/entité qui fait, et l'Objectif (*Goal*) dans une analyse transitive. Il s'agit de procès externalisés, qu'ils soient physiques ou abstraits. Concernant les participants, nous utiliserons les termes d'Agent, pour désigner la cause de l'actualisation du procès et de Medium, qui correspond dans d'autres terminologies linguistiques au terme de Patient (Halliday 2004 : 336). Ces termes permettent de simplifier la description des types de procès (la grammaire hallidayenne comprenant des termes spécifiques pour chaque type de participants aux différents types de procès). La seconde catégorie des types de procès est celle des procès mentaux. Ces procès ont à voir avec la perception, la cognition (*consider, find, believe*), l'affection (*like, love*) et le désir (*wish*). Un des sous-types de procès mentaux qui nous intéresse ici en particulier est celui des procès Verbaux ou Communicationnels, où l'idée est exprimée sous forme verbale/linguistique. Les participants de ce type de procès sont le *Sayer* et *Verbiage*. Le *Sayer* est entendu au sens large qui n'est pas nécessairement un être conscient, comme on le voit dans l'exemple ci-dessous. Enfin, les procès relationnels expriment des relations abstraites d'identification ou d'attribution. Ils sont souvent réalisés sous la forme notamment des copules *be* et *have* (*This Act has jurisdiction*) et des verbes d'état (*become, possess*), mais également de verbes comme *include* (*The powers defined under Subsection X include powers to...*).

Or, les prédicats complexes contenant des noms prédicatifs, puisqu'ils contiennent un verbe et un procès exprimé par un nom, peuvent modifier l'analyse des types de procès sémantique en fonction du nom et du verbe utilisé (Gledhill 2008). Nous donnons ci-dessous plusieurs exemples tirés de notre corpus ou des guides rédactionnels eux-mêmes.

- (42) For the purposes of subsection (6.4), in determining whether reasonable arrangements have been made for the support of a child, the court shall **have regard** to the applicable guidelines. (EN-LEX)
- (43) Please **make an application** for a personal loan. (PLAIN 2011)
- (44) Les notes explicatives ci-jointes vous aideront **dans l'interprétation** des documents juridiques. (Bureau de la traduction 2018)

On voit ici que ces constructions "*have regard to*" (procès mental), "*make a finding*" (procès mental) et "*make an application*" (procès matériel) sont composées en partie de deux verbes génériques *make* et *have*. Ces verbes, dans leur emploi prototypique, expriment respectivement un procès matériel et un procès relationnel. Dans ces exemples cependant, le procès est désigné non par le verbe (qui est appelé Prédicateur dans la terminologie de la GSF), mais par un nom, ici "*finding*", "*regard*" et "*application*" (pour *apply*). Certaines de ces expressions sont le résultats de lexicalisations ("*have regard*" ici). "*Have*" n'exprime donc plus un procès relationnel, mais un procès mental parce qu'il entre en collocation avec *regard*. Dans *make an application*, *make* exprime également un procès (matériel), tout comme *apply*, mais on considère la collocation avec la nominalisation *an application* confère un plus haut degré d'abstraction et d'impersonnalité que le verbe seul "*apply*", bien qu'ils soient partiellement équivalents. On remarque qu'*application* désigne le procès lui-même, comme *apply*, mais également le document administratif qui résulte du procès, ce qui est sans doute pourquoi le prédicat complexe est préféré au verbe simple dans un contexte juridique. Concernant l'exemple en français, *vous aideront dans l'interprétation*, il permet de souligner que les prédicat complexes vont au-delà des structures [verbes + noms prédictifs]. Dans cette exemple ce n'est pas un complément d'objet direct mais un groupe prépositionnel qui exprime la portée du procès, c'est-à-dire que le complément délimite l'étendue du procès sémantique exprimé par le prédicat (Gledhill 2008 : 57) et non pas ce qui est affecté ou modifié par le procès. Dans cet exemple, la portée du procès peut être glosée par la forme verbale "interpréter les documents"; c'est d'ailleurs ainsi que le guide dont cet extrait est tiré suggère de reformuler ce prédicat complexe. Par ailleurs, Bhatia (1994) propose une hypothèse pour expliquer en quoi les nominalisations sont à la fois utiles pour les juristes et complexes pour les non-spécialistes :

Nominalisation certainly helps a legal draftsman to make his provisions more compact and yet precise and all-inclusive; however, it also has the adverse effect of making it more dense and hence difficult to interpret. It takes a lot more time and effort and expertise to unpack, as it were, each nominalised phrase" (Bhatia 1994)

En exprimant un procès sous forme nominale, les prédicats complexes augmenteraient la densité lexicale ainsi que la longueur des phrases et le caractère abstrait et impersonnel. Pour les juristes, ils permettent cependant, d'après Bhatia, de présenter l'information de manière plus compacte et précise, d'où leur emploi fréquent, comme nous l'avons vu au Chapitre 1. C'est en réaction à cela qu'elles sont déconseillées dans les règles centrales du PL. Nous souhaitons donc observer en corpus, la distribution des types de procès sémantiques et la proportion des prédicats complexes : un type de procès est-il davantage représenté dans les différents corpus ? Le nombre de prédicats complexes diminue-t-il lorsqu'on passe d'un discours spécialisé à un discours simplifié, comme indiqué dans les recommandations ?

Il faut cependant souligner que toutes les langues planifiées ne considèrent pas les prédicats complexes comme intrinsèquement complexes, de même que les nominalisations dérivées de verbes ne sont pas toujours considérées comme plus abstraites. Le Basic English (Ogden et Halász 1935), à l'inverse, restreint le nombre de verbes à 15 opérateurs (incluant *be, have, come, put, take, go, etc.* et encourage l'utilisation de prédicats complexes contenant ces verbes très simples et des noms prédicatifs. Ainsi le verbe *attempt* devient-il *make an attempt*. Par ailleurs, les nominalisations sont découragées en PL car considérées comme entraînant l'usage de mots plus longs que les verbes dont ils sont issus (comparez *apply* et *application*) puisqu'on a vu que la complexité lexicale est souvent vue comme corrélée à la longueur des mots en syllabes, mais certains mots dérivés sont en réalité plus aisés à comprendre que des mots courts. Bailin et Grafstein (2001) indiquent :

There appear to be a significant number of instances where mono or bisyllabic words are more esoteric, more unfamiliar, than longer polysyllabic terms. (...) It is conceivable and reasonable that morphologically complex words may actually assist comprehension by providing the reader with tools that would help him or her develop a plausible guess as to the meaning of a word.

Cette idée peut être illustrée par deux termes particulièrement saillants dans le discours juridique, *application* et *act*. Si le ou la lectrice connaît le sens du mot *apply*, on peut supposer que l'ajout du suffixe nominal *-tion* n'empêche pas la compréhension du mot *application*, tandis que *act*, bien que monosyllabique, est employé dans les textes de loi avec un sens spécialisé précis et morphologiquement indistinguable de son emploi non-spécialisé. La préconisation d'éviter les prédicats complexes (désignés dans les recommandations sous le terme *hidden verbs* ou, abusivement, *nominalisation*) repose donc là aussi sur certaines préconceptions de la complexité textuelle, qui ont également influencé le développement des métriques et scores de lisibilité.

3.4 Complexité et simplification en Traitement Automatique du Langage

Jusqu'ici, nous avons vu les notions de complexité qui touchent au lexique et au groupe verbal. Nous passons maintenant à deux niveaux d'analyse supérieurs, ceux de la phrase et du texte. Dans cette section, nous présentons plusieurs métriques de complexité qui ont été développées pour assigner un score numérique à un texte, en lien avec les notions de complexité ouverte et de complexité "absolue" évoquées plus haut.

3.4.1 Mesures automatisées de la complexité lexicale

Comme nous l'avons vu dans la Partie 3.1., la complexité lexicale se mesure souvent selon deux paramètres. La recherche en acquisition du langage, de même que la recherche en TAL, offrent des outils pour mesurer la complexité lexicale.

En effet, certains indices et métriques de complexités ont été implémentées dans des outils pour permettre d'évaluer automatiquement la complexité lexicale d'un ou plusieurs textes, comme le *Lexical Complexity Analyzer* développé par Lu (2012). Cet outil contient par exemple l'une des mesures les plus utilisées pour mesure la diversité lexicale, le *Type-token ratio* (TTR), c'est-à-dire le nombre de mots différents (*types*) par rapport au nombre total de mots utilisés (*tokens*) dans un texte. On peut par exemple supposer que les textes de lois en anglais ont une faible diversité lexicale, car les caractérisations du discours juridique montrent, en particulier dans le cas du discours législatif anglophone, qui se distingue par un nombre important de répétitions. Le problème avec cette mesure est que des textes possédant un grand nombre de mots (*tokens*) donnent des valeurs plus basses de TTR que des textes avec peu de *tokens*, et inversement. Cela est dû à la possibilité pour le nombre de *tokens* d'augmenter indéfiniment, et c'est pourquoi de nombreuses variantes de cette mesure ont été développées. L'outil de Lu (2012), le *Lexical Complexity Analyzer* (LCA) implémente ces différentes variantes, ainsi que de nombreux scores de sophistication et de densité lexicale. Par exemple, on peut mesurer la densité de verbes, c'est-à-dire le ratio de verbes par rapport au nombre total de mots du texte. Ces mesures reposent sur l'étiquetage morpho-syntaxique (*part-of-speech tagging*) et l'annotation automatique des textes. Elles ont été utilisées, entre autres, dans des études en acquisition et didactique des langues, pour prédire automatiquement le niveau d'un texte en se fondant sur des métriques de complexité (Sousa et al. 2020). Il nous

semble intéressant d’explorer ces métriques pour caractériser nos deux corpus, avant et après simplification.

3.4.2 Scores de lisibilité classiques et leurs versions automatiques

La question de la longueur des phrases et des mots est souvent incluse, nous l’avons vu, dans les recommandations, de manière assez précises, comme le montre cet extrait des recommandations du site britannique institutionnel *gov.uk* :

When you use a longer word (8 or 9 letters), users are more likely to skip shorter words (3, 4 or 5 letters) that follow it. So if you use longer, more complicated words, readers will skip more. Keep it simple. (Government Digital Service 2016)

Certains guides ne mentionnent pas explicitement la longueur des phrases ou des mots à atteindre, mais enjoignent les rédacteurs de textes législatifs ou de diffusion du droit de respecter un certain score de lisibilité. Les scores de lisibilité ont été développés pour mesurer le niveau de lecture d’un texte en s’appuyant sur différents paramètres. Dans les guides et recommandations institutionnelles, on trouve en particulier des mentions du *Flesch-Kincaid Reading Ease*, un indice de lisibilité calculé notamment à partir de la longueur des phrases.

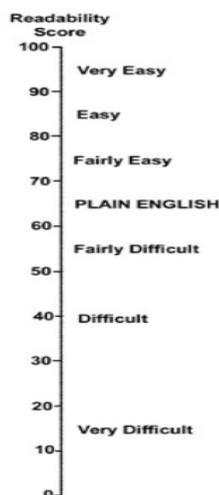
Flesch reading ease : A score of 65 or above is thought to be plain English. This isn’t set in stone though — sometimes specific terms can drag the reading ease down, so aim for a score of at least 60. (Digital Government New Zealand 2020)

Le score *Flesch Reading Ease*, développée par Flesch (1948), vise à évaluer la complexité d’un texte en se fondant sur deux variables : le nombre de syllabes pour 100 mots - les mots polysyllabiques étant considérés comme plus complexes que les monosyllabiques - et le nombre de mots par phrase - là encore, une phrase comportant beaucoup de mots étant considérée comme plus complexe, car elle demande davantage d’effort de mémorisation aux lecteurs. La formule est fondée sur une régression linéaire multiple calculée à partir des scores obtenus en se fondant sur un corpus de référence (étalon), des exercices de lecture pour apprenant.es natif.ives. Les coefficients obtenus sont les suivants pour l’anglais :

$206,835 - 0,846 \times \text{nombre de syllabes pour 100 mots} - 1,015 \times \text{nombre moyen de mots par phrase}$.

Le Flesch Reading Ease (FRE) correspond à une échelle de lisibilité reproduite ci-dessous.

FIGURE 3.3 – Echelle du Flesch Reading Ease (Flesch 2007)



La formule renvoie un score sur 100 pour le texte étudié, variant entre une intervalle de 100.0-90.0 ("très facile à lire") à une intervalle de 30.0-0.0 ("très difficile à lire"). Plus le score est élevé, plus le textes est considéré comme facile à lire.

La formule de Flesch a ensuite été reprise et étendue par Kincaid, donnant lieu au score *Flesch-Kincaid Reading Ease*, qui se fonde également sur le nombre moyen de syllabes par mot et de mots par phrase. Le score obtenu correspond au niveau d'études nécessaire pour comprendre le texte dans le système scolaire étatsunien, où l'échelle va de ce qu'un enfant de 9 ans peut comprendre, à ce qu'un étudiant à l'université peut lire sans difficulté. Il peut ensuite, bien sûr, être converti en un niveau scolaire d'un autre contexte socioculturel. Le score 100, par exemple, correspond selon François & Fairon (2013) à un texte pour lequel un.e élève de CM1 (dans le système français) ou de 4ème primaire (dans le système belge) obtiendrait le score de 75% à un test de compréhension. Le score de 65 préconisé par les institutions néo-zélandaises, correspond au niveau scolaire *8th grade* et *9th grade*, soit un âge de lecture de 13 à 15 ans. Comme pour le vocabulaire, on voit que les formules de lisibilité présupposent un lectorat-modèle, où les lecteur.ices moyen.nes forment un bloc homogène, pour créer une mesure quantitative applicable à n'importe quel texte, sans prendre en compte les variations individuelles (troubles de l'apprentissage, difficultés de lecture). Aux Etats-Unis, la formule de Flesch est utilisée dans plusieurs états et au niveau fédéral, en particulier pour la loi de 1984 portant sur la

simplification de l'assurance santé, afin de mesurer la lisibilité des textes simplifiés ; cette mesure a également été utilisée en Australie pour vérifier l'intelligibilité de plusieurs documents, comme des textes de loi dans l'Etat du Victoria (Asprey 2004). Une seconde mesure classique de complexité/lisibilité est la mesure *Simple Measure of Gobbledygook*, développée en 1969 puis mise à jour par McLaughlin (1969). Elle permet également d'assigner un niveau scolaire à des échantillons de textes en se fondant sur le nombre de mots polysyllabiques et de phrases par textes. Enfin, le test Dale-Chall, une formule qui mesure à la fois la longueur des phrases et la fréquence lexicale, permet de comparer le lexique d'un texte donné à une liste de 3000 mots considérés comme les plus communs, créée par Dale et mise à jour en 1995 (Pishghadam et Abbasnejad 2016). La mesure s'applique à l'origine à des échantillons de textes de 100 mots exactement. Tout comme la formule de Flesch-Kincaid, le test Dale-Chall a été utilisé dans la législation sur le style clair, afin de spécifier un niveau désiré de lisibilité (Leete 1981).

Les mesures classiques et traditionnelles de complexité, notamment la formule de Flesch-Kincaid et le score SMOG, ont fait l'objet de nombreuses critiques. Elles reposent en effet seulement sur deux caractéristiques linguistiques, à savoir la longueur de phrases et la longueur de mots, ignorant donc d'autres traits discursifs qui peuvent avoir un impact sur la compréhension, comme par exemple la cohésion textuelle et la phraséologie (Crossley et al. 2017). Surtout, nous l'avons vu, elles escamotent les variables liées au lecteur ou à la lectrice, telles que les capacités de lecture, le niveau d'alphabétisation et les connaissances préalables. Ces formules échouent donc à prendre en compte la complexité relative dont nous avons parlé plus haut. Se voulant des mesures objectives de la complexité absolue d'un texte, ces mesures sont donc uniquement quantitatives et centrées sur le texte, ignorant les variations socio-culturelles et psycholinguistiques Crossley et al. (2017). Par exemple, la formule Dale-Chall a été critiquée par les chercheurs en acquisition de langue première et en psycholinguistique cognitive car la liste de mots dits "très courants" sur laquelle elle se fonde, bien qu'elle ait été mise à jour en 1995, ignore les différences socioculturelles déjà évoquées (Pishghadam et Abbasnejad 2016), telles que l'âge, le milieu social, le niveau d'études et le degré de maîtrise de la langue. D'autres mesures, centrées cette fois sur la perception et l'attitude du lecteur ou de la lectrice vis-à-vis du texte, ont alors été développées en réponse aux mesures classiques de lisibilité (Pishghadam et Abbasnejad 2016).

Ces scores de lisibilité sont cependant d'intérêt pour notre travail, puisque les institutions anglophones notamment, qui se réclament d'un style clair, les citent

comme références pour établir les standards de lisibilité des textes publiés en ligne. C'est le cas par exemple de la formule de FRE, citée par le gouvernement néo-zélandais. Ces mesures classiques sont en effet abondamment utilisées par les administrations, spécialistes, linguistes et rédacteurs techniques pour mesurer et évaluer le niveau d'intelligibilité de textes de communication entre spécialistes et grand public, par exemple dans le domaine de la santé, pour évaluer des textes d'éducation thérapeutique en ligne (Beaunoyer et al. 2017), ou vérifier la conformité de textes administratifs aux standards imposés par les gouvernements (Pati et al. 2012). Les apports en psycholinguistique ont par ailleurs permis de conclure qu'on obtenait d'assez bonnes corrélations entre ces mesures et le niveau de lisibilité d'un texte, si elles sont conjointes à d'autres mesures O'Brien (2010). Ces formules sont donc pertinentes pour caractériser le passage des textes considérés complexes à des textes supposés en *plain language*. Elles permettent par exemple de proposer une analyse contrastive des genres discursifs, par exemple les textes législatifs et les textes administratifs. Les formules de Flesch sont applicables au français mais que des précautions méthodologiques doivent être prises dans l'analyse des résultats, car " les mots sont en moyenne 1,15 fois plus long dans les textes francophones que dans les textes anglophones." (François et Fairon 2013). A l'origine développée pour être codées à la main sur des échantillons de textes d'au moins trente phrases, ces métriques ont, avec d'autres, été implémentées dans des outils permettant de les mesurer automatiquement sur un grand nombre de textes, et sur des textes plus longs. Ces outils sont disponibles sous la forme de bibliothèques développées dans différents langages de programmation, comme Python (Bansal et Aggarwal 2018). Ces scores ont été utilisés dans des études visant à prédire automatiquement le niveau d'un texte (Sousa et al. 2020). Nous souhaitons explorer cette possibilité de prédiction automatique du registre d'un texte (juridique/style clair) à l'aide de ces mesures.

Nous nous tournons à présent vers les mesures de complexité lexicale et syntaxique développées pour mesurer automatiquement la complexité d'un texte, qui serviront également cet objectif.

3.4.3 Évaluation automatique de la complexité syntaxique

Nous passons maintenant à la complexité dans la phrase, avec les mesures de complexité syntaxique. La complexité syntaxique est liée aux relations entre les entités dans la phrase, et à la manière dont ces relations sont exprimées par la grammaire. Elle est typiquement mesurée en quantifiant le nombre ou la longueur

ainsi que le degré de sophistication de certaines structures syntaxiques (Lu 2010), en particulier le nombre de mots ou propositions par phrase, le nombre de GV ou de certains types de propositions. Nous utilisons empruntons le terme *intricacy* à Halliday et Webster (2013), pour faire référence à l'interaction complexe de plusieurs choix syntaxique considéré complexes, par exemple, une proposition qui inclurait une tournure impersonnelle, un passif et un prédicat complexe. Selon les mesures classiques de Halliday & Webster (2013) et Ure (1979), il suffit de compter le nombre de propositions et de le ramener au nombres de phrases pour avoir une mesure de la complexité syntaxique. Cependant, plusieurs autres mesures ont été proposées, notamment en acquisition L1 et L2, qui s'appuient notamment sur les T-units, définies comme suit : *The T-unit is defined as consisting of a main clause plus all subordinate clauses and nonclausal structures that are attached to or embedded in it.* (Hunt 1970) Pour une mesure plus précise de la complexité, il faut par ailleurs s'intéresser au type de relation entre les propositions dans une phrase. Traditionnellement, deux types de relations sont distinguées entre les propositions : la subordination ou hypotaxe (en anglais *hypotaxis*), qui sert à décrire l'agencement des propositions dépendantes avec les propositions principales, ou la coordination ou parataxe (*parataxis*), qui peut être explicite (présence d'une conjonction de coordination) ou implicite (simple juxtaposition). Dans certains modèles, un troisième type de relation existe, l'enchâssement, pour décrire notamment les propositions relatives, les propositions nominales en *that* et les propositions infinitives ; ces propositions enchâssées fonctionnent comme groupe nominal ou modifieur d'un GN, c'est pourquoi elles sont analysées différemment (Banks 2011).

Dans les mesures classiques de lisibilité évoquées plus haut, comme le score FRE ou SMOG, la longueur des phrases est corrélée avec la complexité syntaxique :

Almost all readability formulae equate sentence length with syntactic complexity.
(Bailin et Grafstein 2001)

C'est pourquoi on recommande de limiter le nombre de mots par phrase (de 15 à 25 mots selon les guides), afin de limiter l'accumulation de propositions au sein d'une même phrase, ce qui est vu comme exigeant un plus grand effort cognitif. Bailin et Grafstein (2001) suggèrent cependant que deux phrases coordonnées peuvent au contraire favoriser la compréhension plutôt que l'empêcher, grâce à la présence de conjonctions de coordination et de subordination qui, bien que rallongeant la phrase, rendent plus explicite la relation entre les propositions et l'intention communicative.

La recherche en acquisition du langage et en TAL propose plusieurs métriques

pour évaluer automatiquement la complexité syntaxique. En acquisition L2, les mesures développées pour évaluer la qualité ou l'évolution des productions écrites d'apprenant.e.s ou d'étudiant.es natif.v.es (Crossley 2020) incluent les structures considérées comme les plus complexes ou sophistiquées, notamment les propositions relatives, les passifs et les propositions en WH-. . Le TAL a donc permis d'étendre les mesures de complexité syntaxique en les systématisant pour des grands corpus, grâce aux outils d'étiquetage et d'annotation.

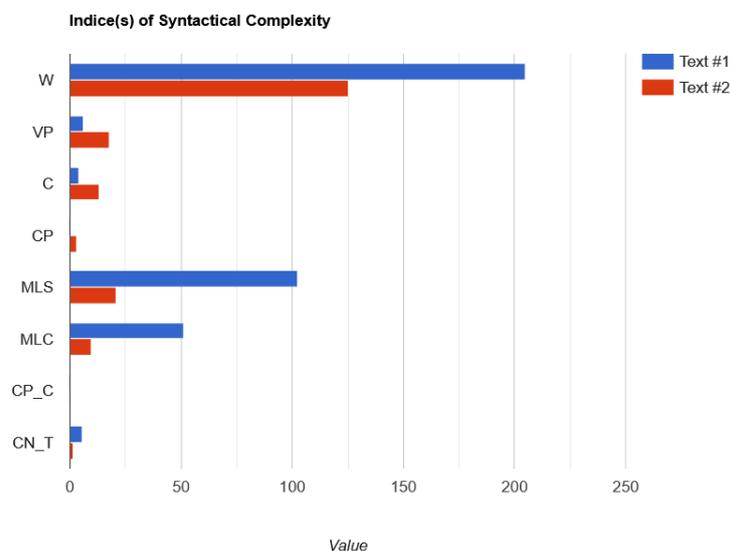
Nous décrivons en particulier l'outil développé par Lu (2010), le *L2 Syntactic Analyzer*, qui combine quatorze types de mesures différents pour l'anglais pour illustrer ces mesures. Il permet de mesurer la longueur moyenne des phrases, propositions et T-units, mais également la quantité de conjonctions de subordination et de coordination, et le ratio de certaines structures, en particulier les groupes nominaux complexes (*complex nominals*) et les groupes verbaux par proposition, par T-unit et par phrase. Pour cela, le L2SCA repose sur la décomposition grammaticale (*parsing*) réalisée automatiquement par le Stanford Parser (De Marneffe et Manning 2008), qui analyse les structures syntaxiques des phrases et les décompose en arbres syntaxiques, après avoir étiqueté les textes (*POS-tagging*). A titre d'exemple, nous présentons certaines de ces métriques de complexité syntaxique mesurées sur deux extraits de textes issus de nos corpus. Le premier (exemple 45 et *Text 1* sur la figure) est extrait de l'*Adoption Act* de 2002 du Royaume-Uni ; l'extrait étudié porte sur le droit à l'obtention d'un acte de naissance par une personne adoptée une fois majeure. Le second extrait (exemple 46, *Text 2* sur la figure) est un texte destiné aux justiciables, qui explique ces mêmes droits pour une personne majeure ou mineure au Royaume-Uni.

- (45) On an application made in the prescribed manner by an adopted person— a record of whose birth is kept by the Registrar General, and who has attained the age of 18 years, the Registrar General must give the applicant any information necessary to enable the applicant to obtain a certified copy of the record of his birth. "Prescribed" means prescribed by regulations made by the Registrar General with the approval of the Chancellor of the Exchequer. Before giving any information to an applicant under paragraph 1, the Registrar General must inform the applicant that counselling services are available to the applicant— from an adoption society, an organisation within section 144 or an adoption society which is under Article 4 of the Adoption (Northern Ireland) Order 1987 (SI 1987/2203 (N.I. 22)), if the applicant is in England and Wales, at the General Register Office or from any local authority or adoption support agency, if the applicant is in Scotland, from any council constituted under section 2 of the Local Government etc. (Scotland) Act 1994 (c 39), if the applicant is in Northern Ireland, from any Board.
- (46) If you are under the age of 18, you can be adopted if you are not married **or** in a registered civil partnership. As a child, you will have little say over who adopts you, **but** you should be

consulted by the adoption agency **or** social services department involved in the adoption. If you are under 18, you have no legal right to know the identity of your birth parents. Once you reach 18, you have a right to apply to see your birth record **and** to use the Adoption Contact Register. The register helps adopted people get contact details of birth relatives who have also registered. You can find out how to see your birth record **or** use the Adoption Contact Register on GOV.UK.

La Figure 3.4 montre une représentation graphique de certaines de ces métriques.

FIGURE 3.4 – Métriques de complexité syntaxique du L2SCA (Lu et Ai 2015) mesurées sur un texte législatif et un texte en style clair



L'indice W renvoie au nombre de mots par texte. Le Texte 1 contient 205 mots, et le Texte 2 en comprend 125 mots (*tokens*). Que révèlent les autres indices ? On peut tout d'abord noter la longueur moyenne des phrases (MLS) et des T-units (MLT, unité syntaxique) dans le texte juridique, avec 102 mots par phrase et par T-unit en moyenne contre 20 mot par phrase et 17 mot par T-unit en moyenne pour le texte destiné aux non-spécialistes. Le Texte 1 contient en effet de nombreuses propositions subordonnées en *if* et énumérations afin de couvrir toutes les possibilités d'application de la loi, mais il n'est composé que de deux phrases ; le Texte 2 contient également de nombreuses propositions subordonnée en *if*. L'extrait de texte législatif contient également quatre fois plus de GN complexes, comme l'indique la métrique CN_T sur la figure, qui mesure le ratio de GN complexes par T-unit (5,5 pour le Texte 1 contre 1,4 pour le Texte 2). On voit en effet dans l'exemple 45 de nombreux GN de type GN1 + préposition + GN2 contenant des GN eux-mêmes pré-

ou post-modifiés, comme *a certified copy of the record of his birth*. L'exemple 46, tiré d'un extrait en style clair, contient quant à lui six fois plus de groupes verbaux (*verb phrases*, VP sur la Figure 3.4) que le texte juridique (18 à 3), et trois fois plus de propositions (*Clauses*, C sur la figure). Là encore, ce résultat est cohérent avec les recommandations pour un style clair, qui encouragent l'utilisation de formes verbales. Le Texte 2 est également caractérisé par un plus haut degré de coordination d'après la mesure CP (*coordinate phrases* (les conjonctions de coordination sont en gras dans le texte)). Cette métrique semble confirmer l'idée de Bailin et Grafstein (2001) sur la coordination (réalisant d'après l'outil une forme de complexité syntaxique) comme potentielle aide à la compréhension. Les extraits sont donc en cela conformes à leur registre, avec un texte spécialisé constitué de très longues phrases et des GN complexes, et un texte destiné aux justiciables qui respecte les recommandations sur la longueur des phrases et l'utilisation de formes verbales.

Ce type de mesures a été utilisé en acquisition L2 pour mesurer la complexité d'écrits d'apprenant.es, mais également pour explorer la classification automatique du niveau d'un texte. En effet, les métriques de complexité lexicale et les scores de lisibilité décrits plus haut, ainsi que les mesures de complexité syntaxique illustrées ici, ont été incorporées dans une chaîne de traitement pour calculer ces métriques sur des textes et tenter de prédire, grâce à une classification automatique, le niveau de textes anglophones produits par des apprenant.e.s de l'anglais, dans l'étude de (Sousa et al. 2020), au sein du projet Ulysses 2019³.

Pour le français, il existe également des outils automatiques destinés à mesurer la difficulté au niveau textuel, et qui combinent les mesures de complexité syntaxique et lexicale avec des scores de lisibilité. Tout d'abord, les métriques du LCA, L2SCA et de Textsat ont été adaptées pour le français. Par ailleurs, d'autres outils existent pour mesurer la complexité de textes grâce au TAL. En particulier, François et al. (2020) ont développé une interface Web spécifiquement destinée aux textes administratifs, rédigés par le gouvernement belge francophone (et étant donc des sous-types du discours juridique, comme nous l'avons vu au Chapitre 2). Cet outil est donc particulièrement intéressant pour nous, puisqu'il est adapté au français (à partir d'un corpus de textes administratifs issus de la Belgique francophone). L'outil mesure lui aussi le nombre de structures syntaxiques imbriquées, le nombre moyen de mots par phrase, proposant en tout "cinq indices de lisibilité portant sur des aspects spécifiques du texte analysé (difficulté lexicale globale, complexité de la syntaxe, niveau de personnalisation ou de cohérence du texte, etc.)" (François et al. 2014). Il

3. Auquel nous avons participé avec Ballier, Gaillat, Zarrouk, Simpkin, Stearns et Sousa. <http://www.clillac-arp.univ-paris-diderot.fr/projets/ulysses2019>

donne également des conseils aux rédacteurs et rédactrices pour simplifier le texte, en se fondant sur ces mesures, pour proposer automatiquement des pistes aux rédacteur.ices afin de simplifier leur texte. Dans la Figure 3.5 est présenté le résultat de mesures de l’outil et ses suggestions sur un extrait du Code pénal français

FIGURE 3.5 – Exemple d’utilisation de l’outil Amesure (François et al. 2014)

The screenshot displays the 'Analyse détaillée' interface of the Amesure tool. On the left, a 'Palette d'analyse' lists various linguistic categories with their respective counts: 'Analyse des phrases' (3), 'Subordonnées' (5), 'Relatives' (0), 'Complétives' (0), 'Autres' (3), 'Voix passive' (5), and 'Toutes' (0). The central 'Texte annoté' shows a sentence from the French Penal Code with highlighted segments. On the right, a 'Phrase' box provides a summary of syntactic difficulties and word counts for the highlighted phrase.

D’autres études ont utilisé le TAL et l’apprentissage-machine (*machine-learning*) pour étudier ou améliorer la lisibilité de textes juridiques. Curtotti et al. (2015) utilisent un modèle pour prédire le niveau de difficulté de phrases, à partir de scores de lisibilité et d’autres métriques calculées, sur un corpus de phrases extraits du discours législatif étasunien analysé avec le Stanford Parser. L’outil VisiLaw Markings de LoPucki (2014) permet quant à lui d’analyser et de visualiser la structure des phrases dans des textes législatifs, en se fondant sur le *parsing*); l’outil repère et signale par des éléments typographiques plusieurs structures syntaxiques comme les phrases, en particulier les conjonctions de coordination et les propositions conditionnelles en IF...THEN. En décomposant la phrase, Visilaw vise à faciliter la lecture et la compréhension des phrases. Ces différents projets confirment la pertinence des outils du TAL pour étudier la simplification de textes juridiques. Dans la suite de ce travail, nous explorons l’apport possible des outils développés par Sousa et al. (2020), développés pour l’anglais puis adaptés au français, pour l’étude du style clair en corpus. Nous souhaitons nous appuyer sur ces métriques et des outils de classification automatique pour caractériser les textes du corpus LEX des textes du corpus PLAIN.

3.5 Résumé du Chapitre 3

Dans ce chapitre, nous avons vu que la complexité en linguistique est une notion multi-dimensionnelle, fortement dépendante de l'interprétation *en contexte* des textes produits, et du lectorat. Plusieurs approches de la complexité linguistique sont possible, selon des cadres théoriques divers.

Dans les guides rédactionnels, les éléments les plus visés car considérés comme complexes sont les suivants : dans le lexique, le vocabulaire soutenu et la terminologie, ainsi que certains schémas phraséologiques ; dans la grammaire, ce sont le passif et les prédicats complexes contenant des noms prédicatifs ; au niveau de la phrase, la longueur des phrases. Notre hypothèse est que ces préconisations sont liées à des stéréotypes sur la complexité linguistique et à une forme d'internalisation d'une vision traditionnelle de la complexité linguistique.

Notre objectif est de confronter les textes produits par les institutions juridiques à la conceptualisation et aux mesures de la complexité proposées dans la littérature. Nous souhaitons ainsi interroger ces recommandations et l'outillage de la simplification de textes juridiques en textes "clairs et simples". Les concepts et mesures de complexité sélectionnés pour ce travail s'appuient à la fois sur ces recommandations et sur d'autres dimensions de la complexité. En effet, les traits que nous avons sélectionnés pour notre étude incluent le passif, ainsi que le lexique et la terminologie, qui sont, nous l'avons vu, très présents dans les guides rédactionnels. Cependant, notre approche s'appuie également sur une approche "TAL" et l'utilisation de métriques de complexité pour classifier automatiquement les textes des corpus, ainsi que sur d'autres dimensions de la complexité, notamment phraséologique. Dans le chapitre suivant, nous décrivons plus précisément les questions de recherche et méthodes retenues pour examiner ces questions.

Deuxième partie

Cadre méthodologique

Chapitre 4

Questions de recherche et méthodologie

Dans ce chapitre, nous présentons les différentes approches méthodologiques utilisées pour étudier la reformulation des textes juridiques par les institutions et associations, dans des corpus décrits au Chapitre suivant. Nous mettons en avant une approche descriptive de l'analyse de textes, à l'aide des outils les plus récents du Traitement Automatique des Langues et de la linguistique de corpus. Nous proposons une méthode systématique pour comparer un corpus de textes réputés complexes à un corpus de textes dits simples, afin de repérer les stratégies de reformulation et de simplification aux niveaux macro- et micro-textuels, en français et en anglais, et dans différents contextes culturels et discursifs.

L'un des objectifs de cette thèse est d'évaluer si les recommandations publiées par les institutions ou organisations qui se réclament du style clair sont réellement appliquées dans les textes produits pour les justiciables, et, sinon, quelles stratégies sont employées à leur place. Cette question du degré d'adhésion aux standards rédactionnels est typique de la recherche sur les langues contrôlées.

Nous souhaitons également plus globalement caractériser les textes réputés clairs et simples par rapport à leur version spécialisée, avec une dimension comparative genre à genre (*genre-based*), mais également entre les langues, contextes et traditions institutionnelles.

4.1 Différents cadres méthodologiques et théoriques pour étudier le style clair

4.1.1 Cadres et réflexions théoriques

La notion de style clair dans le cadre de la diffusion du discours juridique étant, nous l'avons vu, une notion plurielle, notre travail s'inscrit dans plusieurs cadres analytiques. Tout d'abord, comme mentionné précédemment, nous utilisons en partie le cadre de la Grammaire Systémique Fonctionnelle, qui offre une grille d'analyse et une approche structurée applicable à différents genres et registres discursifs. Plus généralement, la perspective que nous avons sur les textes est une perspective fonctionnelle, qui nous permet de relier concrètement les caractéristiques linguistiques aux fonctions rhétoriques et communicatives qui sous-tendent les schémas et traits lexico-grammaticaux observés en discours (Biber et al. 2020).

Notre approche se réclame également de l'école française d'analyse de discours. Quelles stratégies discursives peuvent être mises au jour à travers l'étude du style clair en droit ? Que disent-elles des relations entre les institutions et les justiciables ? Les textes étudiés sont produits par des acteurs institutionnels ou associatifs, qui ont un rapport d'autorité inégal avec leur lectorat, puisqu'ils possèdent des connaissances sur le domaine spécialisé du droit, et tentent de réintroduire une forme de dialogisme avec les justiciables (Preite 2018). En étudiant certains traits linguistiques et leurs fonctions dans les genres juridiques et leurs versions simplifiées, nous souhaitons explorer la notion de voix discursive ou *ethos* dans la communication institutionnelle. L'*ethos* est défini par Maingueneau comme l'image rhétorique construite par l'orateur.ice ou le rédacteur.ice en discours, et investie de valeurs sociales spécifiques au contexte, permettant l'identification du lectorat (Maingueneau 2016). Dans notre étude du corpus PLAIN, nous nous intéresserons en particulier à la voix discursive construite par les rédacteur.ices selon la fonction communicative des textes, qu'ils soient des textes procéduraux (Adam 2011) ou des résumés, ou qu'ils proviennent de différents contextes culturels. En particulier, nous nous intéresserons à cette notion au regard des différences entre la France et certains pays de culture anglo-saxonne.

Notre objectif est également de confronter la notion de la complexité linguistique, telle que présentée dans les guides pour rédiger en style clair, avec les productions réelles des textes se réclamant du style clair. Si, par exemple, le passif est utilisé dans le corpus PLAIN, contrairement aux recommandations, qu'est-ce que cet emploi révèle-t-il de la fonction rhétorique du passif ? Ainsi nous inscrivons-nous dans

une démarche résolument fonctionnelle et descriptive pour analyser le passage du discours spécialisé (représenté par le corpus LEX) au discours destiné aux non-spécialistes (représenté par le corpus PLAIN) Ces deux cadres analytiques, et plus généralement les méthodes de l'analyse de corpus, les notions de genre et de registre discursifs, sont au cœur de ce travail. Nous utilisons également des outils récents du TAL. L'objectif de l'utilisation de méthodes statistiques est d'explorer l'outillage de l'analyse de corpus tout en interrogeant son appareil théorique et méthodologique. Ainsi pourrions-nous les comparer aux recommandations officielles du PLM, et voir si d'autres jeux de descripteurs, à différents niveaux discursifs, sont pertinents et permettent de décrire les phénomènes complexes liés à la reformulation du discours juridique. A l'inverse, nous espérons que nos réflexions méthodologiques sur l'utilisation d'outils statistiques et d'apprentissage-machine en analyse de discours pourront également éclairer la recherche en simplification automatique de textes spécialisés, sujet souvent abordé en linguistique informatique (Cardon et Grabar 2019).

4.1.2 Approche *corpus-based* ou approche *corpus-driven* ?

La plupart des études sur les langues simplifiées et sur la simplification des discours spécialisés sont *corpus-based*, c'est-à-dire qu'elles visent certains éléments linguistiques précis, comme la longueur des phrases ou encore l'utilisation du passif (Warnier 2017).

Cette approche est considérée comme descendante et déductive :

Corpus-based studies typically use corpus data in order to explore a theory or hypothesis, aiming to validate it, refute it or refine it. (Tognini-Bonelli 2001 : 84-85)

Les études qui utilisent cette approche ciblent donc des traits linguistiques pré-définis, en présupposant la validité des formes et structures dérivées de la théorie linguistique, dans le but d'analyser les schémas et la variation de ces formes dans le corpus en question (Biber 2010 : 162). Par contraste, dans la seconde approche, dite *corpus-driven*, c'est l'analyse statistique des corpus qui fait émerger les points d'intérêt linguistique de manière inductive :

The corpus-driven approach differs from the standard practice of linguistics in that it makes minimal a priori assumptions regarding the linguistic features that should be employed for the corpus analysis. (Biber 2010 : 162)

Dans cette approche, c'est donc le corpus lui-même qui est vu comme la source préférée d'hypothèses sur la langue et le discours (Gledhill et Kübler 2016).

Le présent travail adopte une approche principalement *corpus-based*, puisqu'il s'appuie sur les guides rédactionnels publiés pour rédiger en style clair présentés au Chapitre 3. Certains traits lexicaux et grammaticaux en particuliers sont donc ciblés dans les corpus LEX et PLAIN, afin d'évaluer les traits principaux du style clair. Ainsi avons-nous choisi de vérifier des hypothèses sur différents traits linguistiques, que nous expliquons en détail plus bas, portant sur les scores de lisibilité et la longueur des phrases, l'utilisation du passif et le vocabulaire approuvé/non-approuvé.

A l'inverse, certains éléments étudiés le sont davantage dans une approche *corpus-driven*, sans hypothèses précises ; c'est l'étude des résultats quantitatifs qui révèlent alors les traits linguistiques qui caractérisent les corpus. Comme nous le verrons plus bas, c'est cette approche, avec un minimum d'hypothèses sur les corpus, que nous adopterons notamment pour étudier les métriques de complexité et la distribution des procès sémantiques.

4.1.3 Exploitation des corpus

Les corpus LEX et PLAIN sont décrits en détail au Chapitre 5, mais nous présentons dans cette section les relations des différents sous-corpus et la manière dont ils seront exploités. Certains sous-corpus du corpus PLAIN constituent véritablement la version "traduite" en style clair des sous-corpus LEX. C'est le cas des sous-corpus intitulés de Brochures et CA-Résumés, en français et en anglais. D'autres, à l'inverse, comme le sous-corpus anglophone EN-CAB, sont des instances de textes rédigés en style clair mais qui n'ont pas de contrepartie complexe dans le corpus LEX, bien qu'ils se fondent à l'origine sur des textes législatifs (auxquels ils font, dans de rares cas, explicitement référence).

Dans le corpus francophone, le sous-corpus francophone FR-DILA a un statut particulier, puisqu'il ne se réclame pas du *plain language*, concept encore récent en France. Il constitue bien cependant la "traduction intralangue" (Bhatia 1983) du discours spécialisé du droit, FR-LEX, en un discours administratif s'adressant aux justiciables. Certains corpus, comme EN- et FR-CA-Jugements et leurs versions en style clair EN- et FR-CA-Résumés, sont la traduction l'un de l'autre. Nous avons constitué deux corpus par langue, l'un composé de textes réputés complexes principalement destinés à un lectorat expert, et d'un corpus destinés aux justiciables non spécialistes. Ces corpus sont eux-mêmes divisés en sous-corpus, qui contiennent des textes correspondant à des genres discursifs et à des fonctions différents, mais également à des contextes culturels différents (France/ Canada francophone notamment).

Les Figures 4.1 et 4.2 représentent sous forme graphique le lien entre les différents sous-corpus des corpus LEX et PLAIN.

FIGURE 4.1 – Relations entre les sous-corpus de EN-LEX et EN-PLAIN

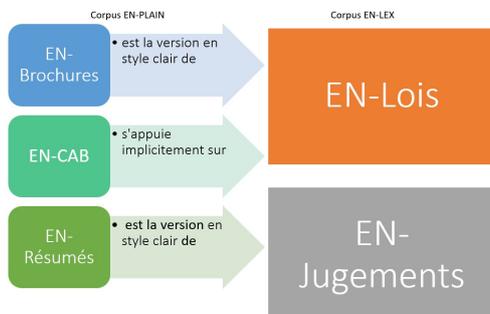
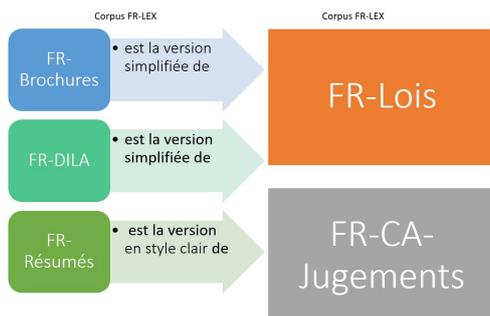


FIGURE 4.2 – Relations entre les sous-corpus de FR-LEX et FR-PLAIN



En matière d'exploitation des corpus, on adoptera deux approches distinctes. Dans la première série d'analyses portant sur les métriques de complexité et les méthodes d'apprentissage, il s'agit en premier d'étudier les corpus LEX et PLAIN sans distinguer les sous-corpus, c'est-à-dire de s'intéresser à ces quatre corpus qualifiés de comparables, "c'est-à-dire où les textes multilingues de départ, tout en abordant un sujet identique, ne sont pas, à la différence des corpus parallèles, des traductions l'un de l'autre" (Raus 2013 : 15). Ainsi aurons-nous deux corpus asymétriques mais distingués par leur lectorat visé et degré de spécialisation (PLAIN vs. LEX) et par leur degré de complexité perçue (simple vs. complexe), mais également par la langue de rédaction des textes. Cette approche est adoptée pour mesurer les métriques de complexité.

Dans les autres séries d'analyses, nous comparons les deux corpus LEX et PLAIN soit en analysant les résultats sous-corpus par sous-corpus, par exemple pour l'étude du passif ; soit nous nous intéressons seulement à certains sous-corpus parallèles afin

d'observer plus en détail, et à partir d'échantillons parfaitement parallèles, comment les stratégies de reformulation et de simplification se reflètent dans certains traits lexicaux et grammaticaux lors du passage d'un texte complexe à un texte simple, quantitativement et qualitativement. Les sous-corpus du corpus EN-PLAIN, qui renvoient à des fonctions différentes de la diffusion du droit, seront également comparés entre eux.

Par ailleurs, nous soulignons que notre approche est bilingue mais que notre corpus de travail principal est le corpus anglophone, ce travail s'inscrivant dans la continuité de la recherche en linguistique anglaise et en anglais de spécialité, d'autant plus que plusieurs des outils utilisés ont été principalement développés pour l'anglais. Comme nous le verrons dans la suite ce chapitre, où nous décrivons la méthodologie employée pour étudier ces différents éléments lexico-grammaticaux, l'étude du lexique et de la terminologie en particulier se concentrera sur les sous-corpus anglophones, même si le sous-corpus francophone sera utilisé pour tester la méthode dans un premier temps.

4.1.4 Questions de recherche

Question de recherche 1 :

Les recommandations pour un style clair s'appuient sur un certain nombre de pré-suppositions sur la complexité linguistique, et préconisent certains traits linguistiques considérés comme associés à la clarté et à la lisibilité. Les productions réelles en style clair reflètent-elles cette même vision de la lisibilité ? Les textes évitent-ils les traits considérés comme complexes par les recommandations ?

Ces idées reçues peuvent être résumées ainsi :

- La longueur des phrases et des mots est corrélée à la complexité : un texte contenant des mots et phrases longues est difficile à lire ;
- Les termes spécialisés et les mots "pompeux" doivent être remplacés par des mots simples et courants ;
- Le passif est plus complexe que l'actif et doit être évité ;
- Les noms, et en particulier les nominalisations, sont plus difficiles à comprendre que les verbes.

La seconde question de recherche concerne l'outillage de notre étude du PL.

Question de recherche 2 :

Comment outiller l'étude du PL dans la communication entre institutions juridiques et le grand public ? Que peuvent apporter les outils du TAL et de la linguistique de corpus pour explorer la simplification des discours spécialisés ?

Notre objectif est d'offrir une méthode systématique pour étudier le style clair en droit. Nous présentons maintenant ces différentes approches.

4.2 *Plain language et machine-learning* : exploration des métriques de complexité et d'un modèle d'apprentissage pour caractériser les corpus LEX et PLAIN

La première série d'analyses effectuée est une caractérisation globale des corpus LEX et PLAIN en termes d'indices de lisibilité et de métriques de complexité. Grâce à des outils récents du TAL et à une méthode empruntée au *machine learning*, nous réalisons une classification automatique des textes de nos corpus en deux catégories (LEX et PLAIN), en nous fondant sur des métriques de complexité calculées sur ces textes. Cette méthode est en continuité avec la recherche récente en sciences humaines tend à utiliser les méthodes d'apprentissage et de classification (apprentissage supervisé et non supervisé, réseaux de neurones), notamment l'analyse de discours (Longhi 2019; Lam et al. 2021). Dans notre étude exploratoire, nous souhaitons interroger la pertinence de ces outils dans l'analyse du style clair, en tentant de répondre aux questions suivantes :

- Que peuvent apporter une méthode d'apprentissage-machine à l'analyse d'un discours spécialisé et de sa diffusion ?
- Les métriques de complexité peuvent-elles être utilisées pour prédire la classe d'un texte du corpus c'est-à-dire le niveau de complexité et le degré de spécialisation du texte (lex vs. plain) ?
- Le cas échéant, quelles dimensions capturées par ces métriques de complexité sont les plus pertinentes pour prédire la classe d'un texte (c'est-à-dire lexicales, syntaxiques, ou scores de lisibilité) ?

4.2.1 Extraction des métriques de complexité : données et outils

Des métriques de complexité et de lisibilité ont été extraits à partir des textes des corpus LEX et PLAIN, en français et en anglais. Les données utilisées, pour l'anglais, sont l'ensemble des textes des corpus EN-PLAIN, soit 865 textes destinés aux justiciables, et les 83 textes du corpus EN-LEX. Pour le français, les 427 textes du corpus FR-PLAIN ont été utilisés. Concernant le corpus de textes spécialisés, les 46 textes du corpus FR-CA-Jugements ont été utilisés, mais seul un échantillon de 457 textes du corpus FR-LEX a été sélectionné. En effet, le temps de traitement de tous les textes du corpus FR-Lois, composé, nous le rappelons, de 3628 articles ou extraits de textes législatifs, aurait été trop important. Le choix a donc été fait de se limiter à un échantillon de ce sous-corpus. La différence de découpage des textes entre EN-Lois et FR-Lois, le premier étant composé de textes législatifs (*Acts*) entiers, le second d'extraits, résulte en un fort déséquilibre entre les jeux de données spécialisées. Ces différences méthodologiques, nous le verrons, ne sont pas sans conséquences sur les résultats.

Comment ces métriques ont-elles effectivement été calculées ? L'extraction de ces métriques s'appuie sur un ensemble d'outils différents réunis dans une chaîne de traitement qui a été développée pour le projet Ulysses 2019 auquel nous avons participé en collaboration avec Ballier, Gaillat, Zarrouk, Simpkin, Sousa et Stearns (voir Ballier et al. (2019a), Ballier et al. (2019b) pour une description complète du projet). Les différents outils de mesures choisis ont été originellement utilisés pour prédire le niveau de textes d'apprenant.e.s de l'anglais (Sousa et al. 2020), afin de représenter les textes sous forme de métriques (vecteurs numériques). Pour chaque texte, on obtient un certain nombre d'indices quantitatifs correspondant à des scores issus de la recherche sur la complexité mentionnés au Chapitre 3. Pour les scores de lisibilité, l'outil implémenté dans la chaîne de traitement est la bibliothèque *Texstat* du langage de programmation Python (Bansal et Aggarwal 2018). Cette bibliothèque inclut la version automatisée de dix formules de lisibilité classiques, notamment les scores *Flesch-Kincaid grade*, *Flesch Reading Ease*, *Dale-Chall score*, *Simple Measure of Gobbledygook* (SMOG) mais également le *text standard*, fondé sur les neuf métriques précédentes et qui indique le nombre d'années scolaires nécessaires pour comprendre le texte. Pour les indices de diversité, densité et variation lexicale, la chaîne de traitement implémente le *Lexical Complexity Analyzer* (LCA)(Lu 2012). Cet outil, comme expliqué au Chapitre 3, se fonde sur *TreeTagger* (Schmid 1994) pour attribuer des étiquettes morphosyntaxiques. Dans notre cas,

27 mesures de diversité et densité lexicale ont été calculées, notamment les ratios des types d’adverbes, d’adjectifs et de mots lexicaux par rapport au nombre total de mots, mais également la densité et la diversité lexicale, fondée sur le nombre de mots différents. Les métriques de complexité syntaxique ont été calculées grâce au L2 Syntactic Complexity Analyzer (L2SCA) (Lu et Ai 2015). Cet outil permet de mesurer 11 indices de complexité syntaxique, comme le nombre de propositions par phrase, ou encore le nombre de groupes nominaux complexes par proposition, en s’appuyant sur les arbres de dépendance du Stanford Parser (Nivre et al. 2016) pour repérer les relations de dépendance. Le L2SCA contient également, pour l’anglais, des modifications pour calculer un certain nombre de mesures de complexité "fondées sur les rapports paradigmatiques qu’entretiennent des formes linguistiques entre elles" (Gaillat et al. 2020). Ces nouvelles mesures implémentées permettent de cibler des structures linguistiques plus fines et se présentent sous la forme de ratios de structures ou formes par rapport à des formes appartenant au même paradigme (Simpkin et al. 2019). Par exemple, le nombre d’occurrences du modal MAY par rapport aux autres modaux *can*, *could*, *may* et *might* est calculé et on obtient un ratio ; de même pour *ought to* par rapport à *should*. Hormis les auxiliaires de modalité, les *microsystem features* ciblent d’autres structures comme les articles définis et indéfinis, les noms composés, génitifs, prépositions et pronoms relatifs, pour un total de 102 métriques. Ce type de métrique n’a cependant pas été adapté pour le français, ce qui entraîne là encore une différence méthodologique entre les jeux de données francophones et anglophones.

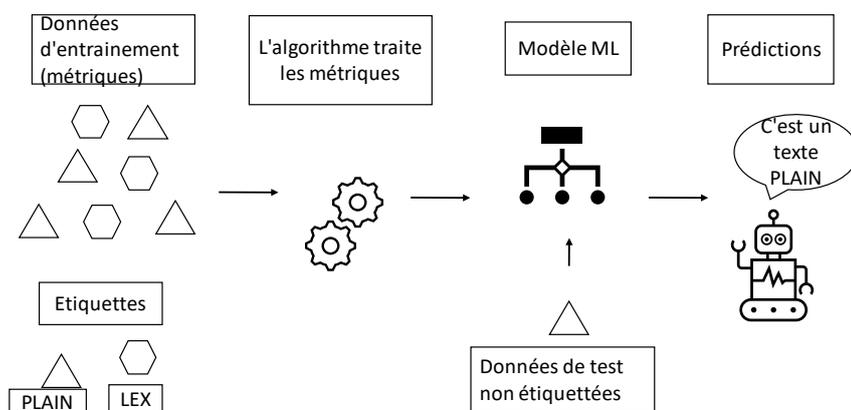
Au total, 150 métriques ont été calculées sur les textes anglophones et 48 métriques sur les textes francophones. La liste totale des métriques de complexité extraites des corpus LEX et PLAIN est présentée dans l’Annexe A. Dans le cadre de la classification automatique, les métriques sont appelées des *features*.

4.2.2 Un modèle d’apprentissage-machine pour étudier les corpus LEX et PLAIN

Après avoir extrait les métriques sur les corpus des textes LEX et PLAIN corpus, sans différenciation des sous-corpus, les indices de complexité pour chaque texte ont été rassemblés dans une même feuille de calcul pour chaque langue, formant le jeu de données sur lequel le modèle d’apprentissage peut travailler. La classification automatique de textes est généralement de deux types : supervisée ou non-supervisée. Dans notre cas, l’apprentissage est dit supervisé car les catégories à trouver sont

fournies au modèle : dans la dernière colonne du jeu de données, pour chaque texte, la classe du texte (LEX) ou (PLAIN) est indiquée, correspondant au corpus de départ, donc au degré de spécialisation et au niveau de complexité supposé. Le modèle d'apprentissage supervisé choisi pour classifier le registre des textes (LEX ou PLAIN) à partir de ces métriques est appelé régression logistique avec régularisation lasso. Que signifie ce terme, et comment l'apprentissage supervisé fonctionne-t-il ? Les données sont d'abord séparées en deux jeux : un jeu d'entraînement (*training set*) qui contient les étiquettes LEX et PLAIN, et un jeu d'essai ou *testing set*, non étiqueté. Dans un premier temps, l'algorithme traite le jeu de données d'entraînement : il utilise les métriques (*features*) pour retrouver les étiquettes de chaque ligne (correspondant à chacun des textes). Dans un second temps, sur le second jeu de données, le modèle ainsi entraîné classe sans connaître les étiquettes (*labels*) des textes au préalable. La Figure 4.3 représente le fonctionnement de l'apprentissage supervisé.

FIGURE 4.3 – Fonctionnement d'un modèle d'apprentissage supervisé



Les résultats de la classification sont ensuite représentés sous la forme d'une matrice de confusion, qui indique combien de textes ont été prédit *plain* et combien le sont vraiment. Plusieurs scores sont calculés par l'algorithme. Ces scores donnent des indications sur la performance du modèle pour classer les textes selon les catégories LEX ou PLAIN.

Ce qui nous intéresse ensuite plus particulièrement est de savoir quelles sont les métriques qui ont le plus grand pouvoir de prédiction pour classer les textes, d'après le modèle, afin de savoir quel(s) type(s) de métriques sont les plus caractéristiques de l'une ou l'autre catégorie. Il faut donc sélectionner un petit nombre de *features* qui sont les plus puissantes pour prédire la classe LEX ou la classe PLAIN. Pour cela, un modèle appelé régression logistique avec régularisation lasso est utilisé. Cette méthode permet d'associer un poids (un coefficient) à chaque métrique (*feature*) afin de savoir lesquelles sont les plus importantes pour réaliser la tâche de classification. Grâce à la régularisation, on peut donc savoir si la classification est très sensible à un *feature* (c'est-à-dire une métrique) donné, et si celui-ci permet d'identifier la classe "LEX" ou "PLAIN". Ce modèle permet de savoir comment chaque *feature* contribue à la classification. Grâce à ces coefficients associés aux métriques, on peut donc savoir quelles métriques permettent d'identifier les textes du corpus LEX ou ceux du corpus PLAIN.

Pour chaque langue (anglais / français), nous avons sélectionné les 8 métriques les plus importantes pour la classification d'après le modèle. Les scores sont analysés quantitativement (distribution des données dans les corpus LEX et PLAIN) sous RStudio, puis une analyse qualitative de concordances correspondant à ces métriques est effectuée à l'aide de SketchEngine.

4.2.3 Des différences quantitatives entre les genres discursifs du corpus PLAIN ?

Notre approche étant *genre-based*, souhaitons également comparer brièvement les différents genres discursifs entre eux, en particulier ceux qui composent le corpus PLAIN. L'une des hypothèses qui nous intéresse est de savoir si les textes de médiation juridique montrent des différences de lisibilité capturées par les métriques selon les fonctions des textes et les sources discursives. Pour cela, dans la présentation des résultats quantitatifs, nous comparerons les différents sous-corpus pour la métrique qui sera désignée comme la plus prédictive par le modèle, en français en en anglais. Si des différences significatives apparaissent, des hypothèses sur ces différences seront évoquées.

Les résultats de la classification et de sélection des métriques sont décrits et analysés au Chapitre 6.

4.3 Approche lexicale et terminologique du style clair en droit

Le second phénomène langagier auquel nous nous intéressons est la question des recommandations présentes dans les guides rédactionnels, un élément récurrent et central du style clair. Plusieurs questions de recherche guident cette approche :

- Les rédacteur.ices des différents sous-corpus PLAIN respectent-ils/elles les prescriptions lexicales faites par les instances promouvant le style clair ?
- Si les mots "interdits" et les termes juridiques sont utilisés dans le corpus PLAIN, sont-ils accompagnés de définitions ou de marqueurs métatextuels d'explicitation ?
- Y a-t-il une plus grande proportion de termes dans les décisions de justice que dans les textes de loi, par exemple ? De même, les résumés de décisions de justice contiennent-ils plus, moins ou autant de termes juridiques spécialisés que les guides juridiques ?

Si les prescriptions lexicales sont respectées, on ne devrait trouver aucun de ces items lexicaux "non-approuvés" et peu de termes (qualifiés de "jargon") dans le corpus PLAIN. L'objectif est d'examiner d'abord le vocabulaire approuvé puis la terminologie, avant de nous pencher sur les contextes explicatifs des termes ou mots considérés complexes.

4.3.1 Liste de vocabulaire non-approuvé

Pour répondre à la première question, nous avons compilé une liste de 342 items lexicaux considérés comme "non-approuvés" par un organisme promouvant le style clair. Cette liste a été compilée à partir du document "*A to Z of alternative words*" publiée par l'association Plain English Campaign (2001a). Cette liste a été créée par les rédacteur.ices de l'organisme, à partir de l'édition de documents :

The 'A to Z of alternative words' and the 'A to Z of legal phrases' were produced by our editors. They simply made a list of words they continually changed, defined them and offered alternatives. We did get them checked by our barrister. (Maher 2021)

Un extrait de cette liste est présenté ci-dessous. La liste complète des mots utilisés se trouve à l'Annexe C.

FIGURE 4.4 – "A to Z" of alternative words - Extrait ©Plain English Campaign

attempt	try
attend	come to, go to, be at
attributable to	due to, because of
authorise	allow, let
authority	right, power, may (as in 'have the authority to')
axiomatic	obvious, goes without saying
B	
belated	late
beneficial	helpful, useful
bestow	give, award
breach	break
by means of	by
C	
calculate	work out, decide

Dans la colonne de gauche, on voit que les mots à éviter appartiennent à un registre soutenu, et sont souvent d'origine latine, polysyllabiques, donc considérés comme demandant un plus grand effort cognitif de la part des lecteurs et lectrices, et moins susceptibles d'être connus. Dans la colonne de droite est présentée une ou plusieurs alternatives claires et simples sont proposées, des mots courants, par exemple *allow* pour remplacer *authorise*.

Il est à noter que la liste des mots lexicaux non-approuvés publiée par PLC contient des éléments polylexicaux, comme la préposition complexe *by means of* dans la Figure 4.4. Pour certains de ces éléments, lorsqu'il s'agissait de groupes nominaux notamment, le nom-tête a été inclus dans la liste de mots non-approuvés. Nous considérons en effet que les recommandations portant sur le vocabulaire ne sont pas uniquement lexicales mais aussi phraséologiques. Les autres ont été inclus dans une liste séparée pour une future étude.

Pour le français, peu de listes libres de droit contenant des mots considérés complexes sont accessibles en ligne. Le chapitre sur cette question se consacre donc essentiellement aux corpus anglophones, mais quelques éléments lexicaux sont également étudiés pour le français, tirés des guides rédactionnels analysés au Chapitre 2, notamment Collette et al. (2002).

4.3.2 Liste de termes pour calculer la densité terminologique

Nous nous intéressons ensuite à la dimension proprement terminologique des textes, en nous demandant si la proportion de termes spécialisés juridiques change lors du passage du discours juridique, représenté par le corpus LEX, à un discours

de médiation, représenté par le corpus PLAIN. Notre hypothèse est que les textes du corpus EN-PLAIN contiennent moins de termes spécialisés que ceux du corpus LEX, et que ceux qui sont présents font l'objet de diverses stratégies d'explicitation.

Comme vu au Chapitre 3, la densité terminologique est définie comme la proportion de termes spécialisés (ici, du domaine du droit) par rapport au nombre total de mots par texte. Pour la mesurer, nous avons donc utilisé une liste de 201 termes issus d'un glossaire anglophone de termes juridiques. Pourquoi utiliser des listes de termes établies *a priori* par des acteurs externes, et pas extraire les termes des corpus EN-LEX et FR-LEX, comme il est commun en terminologie ? Si cette méthode a été envisagée, elle a ensuite été écartée car nous souhaitons dans cette étude nous concentrer sur les termes que les institutions elles-mêmes considèrent comme nécessaires à comprendre pour les justiciables, afin d'explorer la notion de complexité telle qu'elle est présentée par celles et ceux qui se réclament du style clair.

Nous nous sommes appuyées sur un glossaire de termes contenant des définitions publiées sur le site du procureur général de l'Ontario (Canada), qui le présente comme suit :

This legal glossary is a basic guide to common legal terms. A lawyer is in the best position to advise you about your legal rights and responsibilities¹

A partir de ce glossaire, nous avons créé une liste *ad hoc* de 201 termes juridiques. Cette liste se fonde sur un site canadien, or, notre corpus spécialisé est composé de textes publiés par des institutions britanniques, néo-zélandaises et canadiennes. Ces pays appartenant à la même famille juridique, et ayant chacun un système issu de la *common law*, nous avons considéré que la plupart des termes y étaient communs, comme *affidavit*, *homicide* ou encore *probation*, qui font référence à des documents, crimes ou procédures communs. Un nombre restreint de termes comme *certiorari* (voir Chapitre 1) sont spécifiques au contexte culturel (ici canadien).

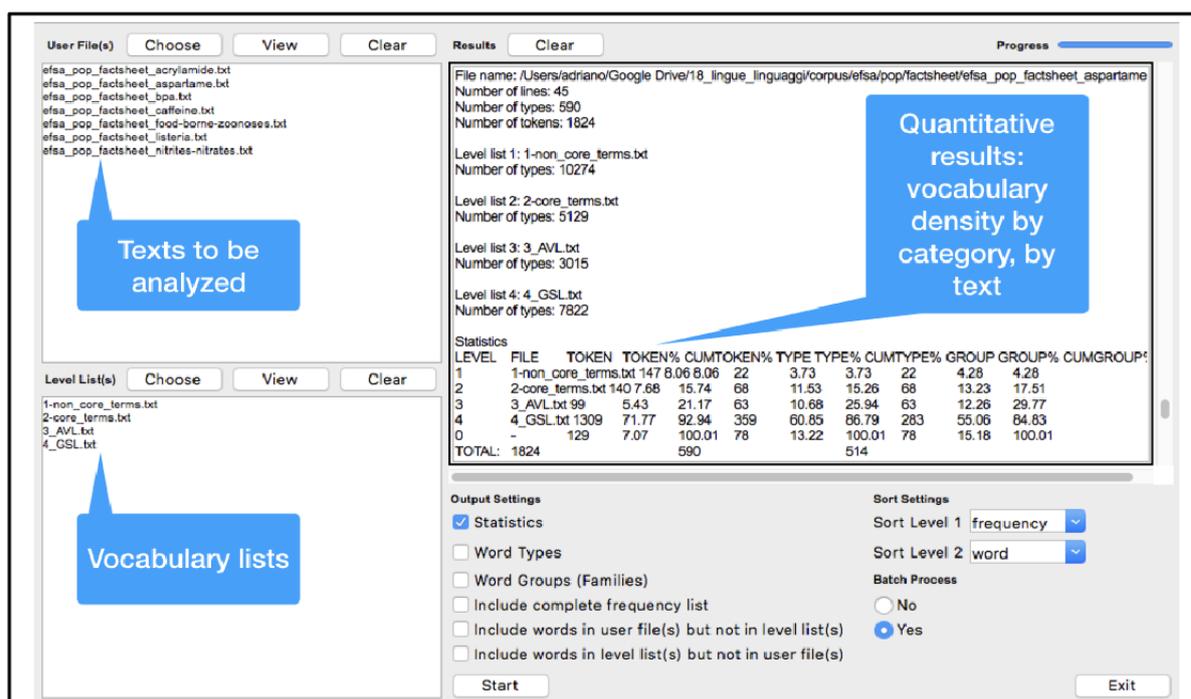
Pour le français, seuls les sous-corpus canadiens FR-CA-Jugements et FR-CA-Résumés sont étudiés pour mesurer la densité terminologique, car, même si certains termes sont communs entre la France et le Canada francophone, le système juridique est très différent en français canadien et en français de France. Une liste de 196 termes a été créée à partir du glossaire canadien. Comme pour l'anglais, les termes complexes ont été exclus de l'étude quantitative.

1. <https://www.attorneygeneral.jus.gov.on.ca/english/glossary/> Page inactive au 8 mars 2022.

4.3.3 Outils logiciels

Une fois les listes de mots non-approuvés et de termes établies, le même logiciel a été utilisé pour mesurer la densité des items lexicaux et terminologique en anglais : AntWordProfiler (Anthony 2014b). La Figure 4.5 extraite de Ferraresi (2019) permet de voir le fonctionnement du logiciel.

FIGURE 4.5 – Fonctionnement d’AntWord Profiler (Ferraresi 2019)



En entrée, le logiciel reçoit les différents corpus en format .txt ainsi que les listes d’éléments lexicaux ou termes. Le logiciel AntWordProfiler permet de comparer la proportion de plusieurs listes de mots dans plusieurs textes à la fois. Pour l’anglais, les textes de chacun des quatre sous-corpus sont concaténés pour obtenir un seul fichier par corpus (afin de faciliter le traitement), qui est chargé dans AntWord Profiler. Puis les listes d’éléments mono-lexicaux du vocabulaire approuvé et non-approuvé ainsi que la liste de termes sont données en entrée. Cela permet d’obtenir la proportion (normalisée en %) des éléments de la liste dans chaque corpus, mais également la liste complète des mots et termes de chaque liste pour chaque sous-corpus ; c’est ce qu’on appelle un profil lexical. Les résultats sont ensuite enregistrés sous forme de fréquences normalisées dans un fichier .csv. Pour les éléments lexicaux non-approuvés, les représentations graphiques ont été faites sous RStudio.

Cette méthode permet donc de vérifier si les textes produits pour diffuser le droit appliquent les recommandations prescrivant d'utiliser une langue et un vocabulaire considéré comme quotidien et familier pour le justiciable, et utilisant peu de termes.

SketchEngine a ensuite été utilisé afin de compléter les analyses obtenues à partir de AntWord Profiler par une étude des contextes des termes et éléments lexicaux les plus fréquents. En particulier, les contextes dans les textes destinés aux non-experts ont été analysés, afin de voir si les mots non approuvés et les termes spécialisés sont accompagnés d'éléments typographiques (guillemets, parenthèses) et/ou expressions métalinguistiques d'explicitation. Les résultats de cette approche lexicale et terminologique sont présentés au Chapitre 7.

4.4 Passif et transitivité dans le discours du droit et sa diffusion

La troisième approche de ce travail concerne des phénomènes langagiers au sein du domaine verbal, à savoir le passif et les procès sémantiques. Le passif est étudié sur tous les corpus, tandis que les procès sémantiques le sont seulement sur un échantillon de corpus.

4.4.1 Le passif dans les corpus LEX et PLAIN

La recommandation d'éviter le passif est, comme nous l'avons vu, l'une des plus fréquentes dans les guides rédactionnels concernant le style clair. Bien sûr, les recommandations reconnaissent également que l'utilisation d'une structure au passif est inévitable ou préférable dans certains contextes, mais son emploi est fortement déconseillé, avec dans certains cas des spécifications quantitatives sur la proportion de verbes à l'actif : *Aim to make about 80 to 90% of your verbs active.* (Plain English Campaign 2013).

Notre objectif est donc d'observer en corpus si cette recommandation est effectivement appliquée dans les textes de médiation juridique. Pour cela, nous vérifions d'abord quantitativement si les différents sous-corpus de PLAIN contiennent peu ou pas de constructions passives. La première hypothèse que nous cherchons à confirmer ou infirmer concerne tout d'abord les formes de passifs périphrastiques canoniques en *be* et *être*. Les formes de passif prototypique dans les corpus LEX et PLAIN sont

extraites par sous-corpus dans SketchEngine (Kilgarriff et al. 2014) en utilisant le langage CQL, à partir du corpus lemmatisé. Pour l'anglais, c'est donc les formes contenant les lemmes *be* suivi du participe passé que l'on cherche ; de même pour le français avec "être".² Après que les formes au passif ont été extraites, on peut obtenir la fréquence normalisée des passifs canoniques par rapport au nombre total de formes verbales conjuguées (*finite verb forms*)³.

Les concordances sont ensuite analysées en contexte et sélectionnées afin de comparer l'utilisation grâce à deux fonctions de SketchEngine. La fonction KWIC permet d'obtenir les lemmes verbaux les plus souvent associés à la voix passive. Inspiré de Banks (2017), nous utilisons les résultats obtenus pour analyser les types de procès des verbes à la voix passive les plus fréquents, afin de dégager d'éventuelles régularités entre type de procès et voix passive. L'outil "Collocations" de SketchEngine permet quant à lui de trouver automatiquement les collocations les plus fréquentes associées aux constructions passives en BE/Etre + VN. Dans notre cas, l'empan choisi est de 3 mots avant et après la constructions passives. On peut ainsi voir se dégager d'éventuels schémas lexico-grammaticaux associés au passif dans les corpus LEX et PLAIN, en particulier explorer quels types de sujets sont fréquemment employés avec le passif et donc thématiques. Pour cela, nous empruntons la classification de Breeze (2017), qui catégorise les sujets des verbes de parole dans un corpus de textes juridiques académiques selon qu'ils correspondent aux catégories suivantes : acteurs institutionnels collectifs (*institutional collective legal actors*), acteur.ice individuel.le du droit (*legal actor, ex juge*), document juridique (*legal document*), sujet impersonnel (*it/this* ou *il*), affaires (*cases*), argument ou principe juridique, ou encore parties, incluant les prévenu.es et accusé.es ou. Pour cette dernière catégorie, nous nous concentrerons en particulier sur comment est faite la référence aux justiciables dans les corpus PLAIN, notamment si la deuxième personne est utilisée.

Cette analyse de contextes a pour but de formuler des hypothèses sur la fonction rhétorique du passif (en particulier s'il est utilisé dans les corpus PLAIN), mais également de voir si les constructions à la voix passive sont insérées dans des segments phraséologiques récurrents.

Concernant les passifs non-canoniques, les formes ne peuvent pas être capturées quantitativement par SketchEngine. Par exemple, pour capturer les *bare passives*, on peut extraire les formes de participes passés, mais celles-ci renvoient pas uni-

2. Étiquettes : [lemma="be"][tag="VVN"] pour l'anglais et [lemma="être"][tag="V.P.*"] pour le français.

3. Étiquette [tag="VBD|VBP|VBZ|VVD|VVP|VVZ|VHD|VHZ|MD"]

quement à ces *bare passives*, car ces formes sont morphologiquement ambiguës : elles peuvent également renvoyer à des formes de passé composés. Les fonctions concordance et CQL sont tout de même utiles pour explorer les formes non-canoniques en français et en anglais, notamment pour repérer les passifs en *get* et passifs pronominaux. Les résultats de cette étude sont présentés au Chapitre 8.

4.4.2 Étude de cas : analyse des procès sémantiques avant et après simplification dans des textes qui portent sur la discrimination

La dernière approche qui nous intéresse est liée aux notions de transitivité et de procès sémantiques, l'un des éléments-clés du modèle de la Grammaire Systémique Fonctionnelle, c'est-à-dire la relation sémantique exprimée par le verbe et ses arguments dans la construction linguistique de l'expérience humaine (Halliday et Matthiessen 2013 : 213). Nous examinons, à travers cette notion, les questions suivantes : quelle est la proportion des types de procès sémantiques dans les textes juridiques, par rapport aux textes en style clair ? Les discours législatif et judiciaire présentent-ils une plus grande proportion d'un certain type de procès sémantiques, en comparaison de ceux utilisés dans les textes de médiation juridique ? Le lectorat visé - spécialistes ou justiciables - affecte-t-il la distribution des procès sémantiques dans des échantillons de texte ? Y a-t-il des différences entre le français et l'anglais ? En revenant au contexte des procès, et à la structure plus générale des textes spécialisés et de leur version "claire et simple", nous nous demanderons également si certains types de procès sont associés à fonctions rhétoriques dans le genre discursif en question. Nous nous intéressons également dans ce chapitre aux prédicats complexes, qui sont considérés comme caractéristiques du discours du droit, afin d'observer ce qu'il en advient dans les versions en style clair des textes.

Pour répondre à ces questions, et puisque les procès sémantiques sont codés à la main, nous nous concentrons sur des échantillons de corpus, en sélectionnant seulement quelques textes des corpus LEX et PLAIN, extraits de EN- et FR-Lois, et des jugements ainsi que de leur version simplifiée. Nous adoptons ainsi une approche corpus mais en nous concentrant sur quelques textes uniquement, dans la lignée de Banks (2002) et Caffarel (2009), qui étudient les ressources transitives et procès sémantiques sur des corpus très restreints ou même de textes isolés. Le "mini-corpus" utilisé et décrit ci-dessous est composé de textes qui portent tous sur la notion de discrimination, choisie car il s'agit d'un concept juridique relativement équivalent

en France, au Royaume-Uni et au Canada. Le concept de discrimination a été choisi parce qu'il est relativement équivalent en droit français et en droit anglo-saxon, et parce que les textes PLAIN se fondaient explicitement sur les textes spécialisés, ce qui nous permet de comparer directement les textes spécialisés et leur version simplifiée dans les deux contextes juridiques.

- Pour le discours législatif :
 - Un échantillon de textes législatifs du corpus Lois, composé pour l'anglais d'un extrait de 4 660 de l'*Equality Act of 2010*, le texte de loi le plus récent définissant la discrimination et, pour le français, les articles du corpus FR-Lois qui définissent et régissent la discrimination (mais également le délit d'incitation à la haine), extraits notamment du Code Pénal et comprenant 1 887 mots (*tokens*) (l'extrait est plus court car il s'agit d'articles et non de chapitres complets du Code pénal) ;
 - Un échantillon des corpus EN-CAB et FR-DILA constitué chacun de trois textes qui constituent une reformulation explicite des textes législatifs mentionnés ci-dessus, qui expliquent aux justiciables le concept de discrimination et les recours légaux possibles lorsqu'on y est confronté.
- Pour le discours judiciaire :
 - en anglais et en français, les motifs de jugements présentés par la Juge Abella pour l'affaire portant sur la loi québécoise contre la discrimination salariale des femmes : il s'agit de l'opinion majoritaire ;
 - pour chaque langue, 2 résumés (respectivement), extraits des corpus EN-CA-Résumés et FR-CA-Résumés, portant sur deux affaires sur la discrimination salariale dont les décisions ont été rendues le même jour.

Une fois les échantillons de corpus constitués, les types de procès sémantiques sont codés manuellement selon les catégories de procès sémantiques de Halliday, en fonction du co-texte et des participants des verbes. Le nombre d'occurrences pour chaque type de procès a ensuite été calculé, et normalisé par rapport au nombre total de verbes codés. Ces types de procès sémantiques incluent, comme nous l'avons vu au Chapitre 3, quatre catégories principales : les procès matériels, liés à ce qu'on fait et à ce qui se produit (par ex. *treat someone unfairly*) ; les procès mentaux, liés à la perception (*see, hear*), à la cognition (*know, understand*), mais aussi à l'affection (*like*) et au désir (*want, hope*) ; les procès verbaux (*verbal*) ; et les procès relationnels, utilisés pour inscrire deux entités dans une relation d'identification (*seem, become, appear*) ou d'attribution (*have, own, possess*). Les procès relationnels

sont par exemple utilisés dans les définitions (par ex "*une discrimination constitue une distinction*"). Nous utilisons également une version simplifiée du modèle pour désigner les participants : pour les procès matériels, on parle d'Agent pour la personne entraînant l'action, et, pour les procès mentaux et verbaux, la catégorie unique de Medium pour les personnes et entités qui ressentent (*Sensers*) ou expriment le procès verbal (*Sayer*).

En plus du type de procès sémantique, nous examinons également différentes caractéristiques du prédicat : si le GV contient un modal et si le GV + COD constitue une construction verbo-nominale (ex *make observations*), puisque, nous l'avons vu, l'utilisation de constructions à verbes supports contenant des nominalisations est découragée dans les recommandations en style clair. L'objectif est de savoir si ces constructions, très présentes dans le discours juridique, disparaissent dans les textes tirés du corpus PLAIN. Nous présentons et discutons les résultats de cette étude de cas au Chapitre 9.

Chapitre 5

Élaboration et description du corpus

5.1 Objectifs et choix du corpus

L'objectif de cette thèse est d'explorer les différents phénomènes liés à l'utilisation du style clair dans le domaine du droit, notamment dans le cadre de la diffusion et de la simplification du discours spécialisé juridique. Dans ce chapitre, nous présentons nos réflexions sur la notion de genre de discours, qui nous ont permis de sélectionner les textes les plus pertinents par rapport à nos questions de recherche, et ainsi de compiler les différents sous-corpus en fonction du lectorat visé, spécialisé ou non expert. Nous décrivons ensuite nos deux corpus, LEX et PLAIN, composés respectivement de textes juridiques réputés complexes, d'expert-à-expert, et de textes réputés simples destinés aux non-experts. Nous détaillons les différents sous-corpus en termes de langue de rédaction, de genre, et de provenance, avant d'expliquer brièvement comment ces données textuelles ont été préparées afin de pouvoir être explorés grâce aux outils de la linguistique de corpus.

5.1.1 Étudier le discours juridique et sa diffusion dans le cadre du style clair

Notre objectif est de comparer différents sous-corpus de textes représentatifs de genres discursifs du droit, rassemblés dans l'archive de textes appelée LEX, et de la diffusion du droit auprès d'un public non-expert, rédigés par des institutions,

organismes ou associations qui se réclament du (*plain language*) ou d'un effort de simplification du droit, rassemblés dans le corpus intitulé PLAIN. Nous nous proposons ainsi d'étudier, sous l'angle des langues contrôlées, différents genres issus d'un corpus destiné à un lectorat non spécialisé, au regard d'un corpus rédigé par des spécialistes pour des spécialistes du droit. L'objectif de l'étude de ce corpus est double :

- D'une part, observer le respect ou non des normes liées au *plain language* ou langue claire et simple, dans les différents genres du registre juridique, en particulier dans la communication entre les institutions avec les justiciables, afin d'affiner les recommandations pour un style clair ;
- D'autre part, outiller la caractérisation du discours de diffusion du droit, dans différents contextes linguistiques et culturels, afin de proposer une grille d'analyse des stratégies de simplification d'un discours spécialisé réputé très complexe.

5.1.2 Une approche comparative bilingue français-anglais

L'un des objectifs de la construction de notre corpus, comme déjà mentionné, est d'introduire une dimension comparée à notre étude, puisque nous nous intéressons à l'anglais et au français. Le premier critère de sélection des textes a donc tout d'abord été la langue de rédaction. On s'intéresse principalement à des textes anglophones issus des institutions des pays ayant intégré les principes du style clair à leur législation et/ou à leur stratégie de communication avec les usagers, à savoir le Royaume-Uni, le Canada, la Nouvelle-Zélande. Nos corpus principaux de travail seront donc les corpus EN-LEX et EN-PLAIN. Cependant, nous souhaitons également comparer le discours de la simplification et de la diffusion du droit en anglais et en français. Le corpus francophone est composé de textes principalement produits par les institutions françaises, mais du Canada francophone, car, comme nous l'avons vu au Chapitre 1, de par la politique de bilinguisme du Canada, la notion de *plain language* a été adaptée au français sous la forme de Langue Claire et Simple ; certains organismes l'utilisent pour s'adresser aux non spécialistes, et encouragent son utilisation dans le droit en général. Nous souhaitons ainsi comparer les stratégies de communication entre des institutions où le mouvement pour un style clair remonte à des décennies, comme le Canada, avec celles de la France, où le mouvement n'a jamais connu une telle institutionnalisation en termes de politique linguistique, malgré la publication de recommandations visant à simplifier la langue administrative

(Collette et al. 2002) et l'apparition récente du concept dans les associations françaises, grâce à l'influence des mouvements européens et belges.

Différentes questions de recherche ont guidé ce choix d'un corpus bilingue : y a-t-il des différences significatives de complexité entre les genres juridiques francophones et anglophones, mais aussi entre les textes francophones (France/canada) ? Peut-on mettre au jour des schémas lexico-grammaticaux de la simplification en tant que mode de discours, et, le cas échéant, en quoi diffèrent-ils en français et en anglais ? Quelles sont les stratégies discursives ou pragmatiques qui diffèrent ou se rejoignent entre ces deux cultures juridiques différentes ? L'une des hypothèses à laquelle nous tenterons de répondre est par exemple : les scores de lisibilité des textes de diffusion juridique rédigés en français-canadien et ceux rédigés en France suggèrent-ils une différence de complexité notable entre ces textes, liées à des différences dans les stratégies de communication des institutions avec les justiciables ?

5.2 Élaborer un corpus : une réflexion sur les genres discursifs

Nous avons sélectionné des genres rédigés et destinés aux experts juridiques, réunis dans le corpus LEX, et leurs parallèles simplifiés ou simples, c'est-à-dire directement rédigés avec l'objectif d'être compréhensibles par les non-expert. Les critères qui ont guidé notre choix sont l'aboutissement de notre réflexion sur le genre discursif.

Dans la théorie de la GSF, et de l'anglais de spécialité, le concept de genre discursif est crucial. Il permet de décrire un registre ou une variété de langue au sens hallidayen, c'est-à-dire la manière conventionnelle et typifiée dont des personnes vont se servir de la langue dans une interaction écrite ou orale, à l'intérieur d'un scénario d'actions et d'événements. A l'intérieur d'un même contexte situationnel, différents choix textuels sont disponibles. Ces choix correspondent aux genres discursifs, des types de textes plus ou moins ritualisés, qui relient un événement communicationnel à des traits sémantiques et lexico-grammaticaux, c'est-à-dire l'ensemble de contraintes les pratiques, schémas interactionnels et moyens de communication associés à un contexte donné (Bawarshi et Reiff 2010). Les genres sont également un concept central pour l'analyse de discours française, qui les définit comme des "dispositifs de communication socio-historiquement définis" ou encore des "pratiques sociales stabilisées à travers l'interaction de multiples contraintes de

divers ordres" (Maingueneau 2013 : 83). Dans le cadre de cette étude sur le style clair, les contraintes génériques sont liées en premier lieu à la discipline, puisque nous nous intéressons au domaine du droit, et au contexte situationnel de rédaction (on écrira différemment une décision de justice d'un texte de loi). Ces contraintes dépendent également du lectorat visé. En étudiant l'ouverture d'un domaine vers un public non-spécialiste, d'autant plus dans le cadre du *plain language*, il faut donc s'attacher à caractériser les genres utilisés dans le domaine du droit par et pour la communauté de d'expert, mais également les différents genres disponibles pour la diffusion de connaissances spécialisées, et entre experts et non-experts. Partant de ces observations, et conformément à la littérature sur la constitution de corpus, nous avons tenté de créer un corpus représentatifs de plusieurs genres du discours du droit :

A balanced corpus usually covers a wide range of text categories which are supposed to be representative of the language or language variety under consideration. (McEnery, 2006 :16)

Nous présentons les questionnements et critères qui nous ont permis de construire les corpus LEX et PLAIN de façon à ce qu'ils représentent plusieurs genres discursifs du droit et leur version simplifiée ou vulgarisée.

5.2.1 Registres et genres du droits choisis

L'élaboration de notre corpus spécialisé LEX s'est donc concentrée sur les deux premiers sous-registres, législatifs et judiciaires : d'une part, des textes de lois rédigés par les institutions parlementaires, que ce soit en France, en Nouvelle-Zélande ou au Royaume-Uni ; d'autre part, un genre représentatif du discours judiciaire, les décisions de la Cour Suprême, la plus haute instance judiciaire dans les pays de tradition juridique anglo-saxonne. Notre second corpus, intitulé PLAIN, est quant à lui composé de genres estampillées "clairs et simples" par les institutions et organismes les ayant publiés, ou tout simplement destinés à un lectorat non-expert.

Or, on s'aperçoit que ces deux corpus correspondent aux deux grandes fonctions de communication de la langue juridique : le corpus LEX est composé de textes remplissant une fonction normative, qui comprend selon Cao "les documents qui établissent des faits légaux ou créent des droits et des obligations", qui sont donc "surtout prescriptifs" (Cao 2007 : 10, notre traduction). A l'inverse, le corpus PLAIN rassemble des textes ayant une fonction informative, réalisée selon Cao par "les travaux et commentaires juridiques académiques, le conseil juridique, la correspondance

entre un avocat et ses clients, et d'autres documents utilisés lors de procédures judiciaires [notre cas]" et qui sont donc "surtout descriptifs" (Cao 2007 : 10). Notre recherche s'inscrit donc dans l'étude du passage de textes à visée normative avec des textes à visée informative, qui sont a priori sans effet de droit, mais liés à des enjeux d'accès à l'information juridique et administrative. En ce qui concerne le corpus PLAIN, représentant la seconde fonction, nous nous sommes tournés vers des textes disponibles en ligne et qui constituent le versant "informatif" des textes normatifs mentionnés plus haut, qui ont pour soit fonction d'expliquer aux justiciables leurs droits et devoirs au regard de la loi, soit de les guider dans les procédures juridiques et administratives qu'ils peuvent être amenés à rencontrer.

Le choix des genres discursifs pour l'élaboration de nos corpus est donc lié aux différentes fonctions du discours du droit, et à ses sous-variétés.

5.2.2 Choix des sources discursives : différents acteurs du droit et de la diffusion du droit

Nous avons donc en premier lieu choisi des acteurs qui ont intégré la notion de style clair à leur stratégie de communication, ou qui encouragent son utilisation dans le droit et dans la communication avec les usagers non-experts. Concernant le corpus PLAIN, à l'instar de Raus (2014), nous distinguons les "acteurs associatifs" des "acteurs institutionnels nationaux et supranationaux", ces derniers étant définis comme des "organismes reconnus au niveau institutionnel national et/ou international" qui, "en raison de leur statut", "jouent un rôle essentiel dans les politiques linguistiques contraignantes ou incitatives des différents pays" (Raus 2014). Ces différentes instances de diffusion du juridique permettent d'élaborer des hypothèses quant aux différents modes du discours de simplification, que nous détaillerons au chapitre suivant. Notre corpus PLAIN mélange donc des textes institutionnels et associatifs, qui ont pour visée de simplifier le discours du droit pour le rendre compréhensible à l'utilisateur non-expert.

L'objectif de notre corpus LEX est d'être représentatif des discours du droit auquel peut être confronté le justiciable - principalement les textes de lois - et qui peuvent influencer sur leurs droits. Les différentes institutions juridiques ne publient pas les mêmes genres, comme nous l'avons vu, ils appartiennent à des sous-registres différents : les institutions législatives publient en particulier des lois (*statutes*), institutions exécutives, des décrets, les institutions chargées de la procédure judiciaire, comme la police et les différentes agences et organismes gouvernementaux, produisent

les documents du discours judiciaires. Les tribunaux et Cours, sont caractérisées mais produisent également le discours juridictionnel, c'est-à-dire le discours de l'ensemble des juridictions qui publient des décisions ou jugements. Ces décisions sont publiques dans le cas des plus hautes instances juridiques. Dans le cas des pays de tradition juridique de *common law*, comme le Canada, les décisions des cours suprêmes font également jurisprudence : fondées sur le principe du précédent, elles sont également créatrices droit et peuvent être utilisées dans d'autres affaires judiciaires : on parle alors de discours jurisprudentiel. Pour le corpus LEX, nous avons donc choisi de recueillir des textes de différentes sources principales du droit, les instances législatives et les instances juridictionnelles et judiciaires.

5.2.3 Le Web comme corpus

Nos corpus, en particulier le corpus PLAIN destiné aux non-experts, ont été compilés à partir de genres disponibles en ligne, afin de refléter la numérisation croissante de la communication juridique. Ce mouvement, commencé depuis plusieurs années et qui s'accélère, destiné notamment à réduire les temps d'appel pour les institutions, est lié également à l'amélioration de la communication numérique pour les usagers : il est par exemple désormais possible de communiquer avec les institutions judiciaires, comme la police, par messagerie instantanée. Il est également désormais possible d'effectuer de nombreuses démarches en ligne, qu'elles soient liées à des procédures administratives (ouverture de droits sociaux) ou judiciaire (pré-dépôt de plainte, par exemple). Certaines démarches se font d'ailleurs uniquement en ligne, ce qui n'est pas sans problèmes en termes d'accès au droit. C'est pour refléter cette évolution et numérisation croissante, qui n'est pas sans incidence sur la recherche en linguistique elle-même, que nous avons choisi d'étudier des textes juridiques et leur version simple disponibles en ligne, et donc d'utiliser le web comme corpus (*web as corpus*). Ce choix reflète par ailleurs une tendance dans les travaux en sciences du langage. Le Web permet d'obtenir rapidement un grand nombre de données textuelles. La recherche en linguistique a depuis plusieurs années créé des typologie des genres web selon des critères à la fois linguistiques et fonctionnels, dans le cadre de la constitution de "super-corpus" destinés à représenter toutes les variétés et les différents genres discursifs, écrits ou oraux, disponibles sur Internet (voir par exemples Biber 2007). Par ailleurs, des travaux récents se sont attachés à caractériser les nouveaux genres apparus avec l'avènement d'Internet. Certains auteurs se sont donc intéressés aux weblogs (blogs) (Maingueneau 2013), d'autres à des genres multi-modaux comme les vidéos TED Talks, notamment dans le cadre

de la diffusion de connaissances spécialisées (par exemple Mattiello, 2017). Le choix d'un corpus composé de textes disponibles en ligne s'inscrit donc dans la continuité de ces recherches. Ce choix reflète également de nouvelles formes de transfert des connaissances spécialisées due à la numérisation, qui touche les domaines scientifique, médical (Harvey et al. 2013) mais aussi, comme nous l'avons vu, juridique (Turnbull 2014). L'une des questions sous-jacentes est de savoir si l'accès accru aux genres constitutifs du discours juridique, présenté comme une stratégie de communication qui vise à augmenter l'accès au droit et à simplifier les procédures, est corrélé à un degré de lisibilité et de simplicité élevé de ces textes.

5.2.4 Ouvrir le domaine spécialisé du droit : fonctions des genres choisis pour le corpus PLAIN

Nous avons présenté les genres discursifs corpus LEX, choisis selon leurs fonctions et selon le registre du doigt qu'ils représentent. Concernant maintenant le corpus PLAIN, nous distinguerons deux fonctions principales pour les textes qui le composent, composé de brochures et autres textes explicatifs publiés par des institutions et associations.

La fonction principale des brochures et pages web publiées par les institutions dans le domaine du droit est d'aider l'utilisateur non-spécialiste à accomplir une procédure administrative (demande d'allocation, démarche liée à l'accès à la santé) ou juridique (dépôt de plainte, adoption, participation à un procès en tant que témoin). Ces textes, comme nous l'avons vu au Chapitre 1, ne sont pas uniquement destinés à faire comprendre aux non-spécialistes les pratiques et le fonctionnement du domaine du droit, et pas seulement à expliciter le texte de loi spécialisé. Ils offrent, selon Preite, dans son étude de la diffusion de la terminologie juridique par le Ministère de la Justice française, "des instructions pratiques destinées à faciliter ses démarches lorsqu'il [le justiciable] se trouverait dans l'exigence de faire valoir ses droits" (Preite 2016).

Il arrive que le contenu de certains documents ne relève pas uniquement du discours procédural, comme par exemple les guides juridiques expliquant les différentes étapes de la procédure de divorce ou pour l'obtention de l'allocation chômage, ou à l'inverse relève seulement d'un discours de diffusion de textes spécialisés (par ex. la version simplifiée de la Convention Européenne des Droits de l'Homme). Cependant, la plupart des brochures ou fiches d'informations sur des questions juridiques peuvent être qualifiés d'hybrides, car ils contiennent à la fois des éléments du discours

instructionnel, et des éléments de transfert des connaissances spécialisées juridiques permettant de recontextualiser cette procédure pour l'utilisateur. On trouve par exemple dans ces textes de passages qui rappellent leurs droits aux usagers ou qui explicitent ce que dit la loi. Ces éléments sont mêlés aux explications exposant la procédure ou la suite d'actions que le justiciable aura à effectuer pour l'administration ou la justice.

Les résumés des décisions de la Cour Suprême du Canada ont quant à eux une fonction plus purement informative et explicative, et non procédurale. Ils sont essentiellement informatifs et descriptifs et s'inscrivent dans la problématique de la simplification et du résumé. Ainsi ces deux fonctions amènent-elles à d'autres hypothèses, par exemple de savoir si les textes procéduraux du corpus PLAIN contiennent des schémas lexico-grammaticaux absents des textes à visée informative (les résumés des jugements de la Cour Suprême, en l'occurrence).

5.2.5 Un corpus synchronique ? Espace et temporalité du juridique

La constitution d'un corpus composé de textes issus d'Internet a entraîné des questions méthodologiques, dues en partie à la variabilité du contenu des ressources en ligne, et des textes juridiques eux-mêmes. L'une des questions méthodologiques fondamentales lors de l'élaboration d'un corpus est en effet celle de la diachronie, ou de la synchronie du corpus. En effet, les textes, juridiques ou de diffusion du droit, ont été pensés comme une sorte de "photographie" des textes accessibles aux praticiens et aux usagers du droit dans un temps restreint - dans notre cas, les textes de notre corpus ont été récoltés sur un an, de 2017 à 2018. Cela correspond donc à la définition de la synchronie selon Bowker & Pearson (2002 : 11-13). Cependant, lorsqu'on regarde de plus près le discours juridique, la question de la temporalité se complique. Ainsi, certains textes du corpus PLAIN disponibles à ces dates s'appuient sur des textes juridiques écrits plusieurs années et même dizaines d'années auparavant. L'une des brochures du corpus PLAIN, qui explique comment un.e ressortissant.e d'un autre Etat peut faire appel d'un ordre de quitter le territoire néo-zélandais, se réfère par exemple à des textes législatifs datant de 2009 mais aussi de 1972. Ces deux textes de loi sont présents dans le corpus LEX, et l'on se demandera si des différences en termes de complexité apparaîtront entre des textes juridiques rédigés avec plusieurs années d'écart, le texte de 1972 étant antérieur au *Plain Language Movement*.

Les évolutions politiques reflétées dans la loi, ou au contraire leur pérennité, sont ainsi elles-mêmes reflétées dans les textes informationnels destinés aux non-experts. Ainsi, la sortie de l'Union Européenne du Royaume-Uni votée en 2016 a d'abord entraîné la publication sur le site Citizens Advice Bureau, des bandeaux qui avertissaient les lecteurs des changements prochains susceptibles d'affecter leurs droits, au moment de la collecte du corpus. Depuis, une nouvelle section spécifique au Brexit, est apparue ; les textes collectés ne seraient donc pas exactement les mêmes aujourd'hui. Ces exemples illustrent donc l'importance de la dimension temporelle du discours juridique, pointée par certains auteurs :

Le droit existe à un moment donné et sur un territoire délimité. (Monjean-Decaudin 2010)

Cette citation évoque également la question de la territorialité du droit, que nous avons déjà évoquée. Le fort ancrage culturel des genres est lié aux différences entre les cultures juridiques et administratives ; parmi des pays appartenant à la même tradition juridique (droit civil ou *common law* dans notre cas), des différences terminologiques et génériques peuvent apparaître, même si le fonctionnement des institutions présente de fortes similarités. Nous avons choisi plusieurs pays anglophones et francophones pour refléter les enjeux de la simplification du droit pour ces différents systèmes juridiques et ces différents contextes culturels. Ils offrent la possibilité de comparer le discours d'institutions semblables ou équivalentes. Pour notre corpus anglophone, des textes provenant du Royaume-Uni, de la Nouvelle-Zélande et du Canada ont été sélectionnés. Ces pays, tous héritiers du système juridique britannique de *common law*, présentent donc de fortes similarités dans l'architecture de leurs systèmes juridiques et législatifs, avec une monarchie parlementaire et une Cour Suprême comme plus haute instance judiciaire, mais comprenant des spécificités nationales, notamment liées aux minorités linguistiques. Côté francophone, en effet, les textes du corpus proviennent pour la plupart de France, pays de tradition de droit civil. Si la simplification du langage du droit a fait l'objet de recommandations dans plusieurs registres (législatif administratif), le concept linguistique de PL n'est pas institué en tant que tel. L'autre source de textes en français est le Canada, sous la forme de jugements et résumés de jugements de la cour Suprême traduits en français.

La multitude de territorialités choisies, ainsi que l'évolution constante du droit, appellent à des précautions méthodologiques, notamment concernant la transposabilité des termes, textes et systèmes juridiques. Elles offrent également la possibilité de comparer non seulement des genres juridiques et leur version en style clair, mais

également des genres équivalents venant de différentes sources énonciatives et de différents pays. Nous avons choisi de considérer comme un bloc les textes législatifs d'un même pays, rédigés à plusieurs années d'écart. Au moment de la constitution de nos corpus, ces textes parfois anciens étaient cités comme textes de référence, encore en vigueur, par les institutions, c'est pourquoi ils ont été inclus dans nos corpus. Il faut cependant noter que les textes de loi datant des années 1980 ont sans doute été largement amendés depuis leur publication, et témoignent probablement de l'évolution du style juridique.

5.3 Description des corpus

Notre réflexion sur les genres discursifs a abouti à la création d'un corpus bilingue composé de genres juridiques destinés à des experts, et de genres de diffusion du droit juridique destinés aux non-spécialistes (justiciables), afin d'étudier les discours de simplification et de diffusion de ce domaine dans différents contextes culturels, en lien avec le *plain language*. Dans cette section, nous décrivons les deux corpus, l'un composé de textes réputés simples, l'autre de textes réputés complexes. Le nombre de mots attribués à chaque sous-corpus correspond au nombre de *tokens* repérés par le logiciel AntConc (Anthony (2014a)), excluant les chiffres et les signes de ponctuation.

5.3.1 Le Corpus PLAIN

Le concept de *plain language* étant au cœur de notre recherche, c'est le corpus composé de textes supposés rédigés en style clair que nous décrivons en premier. Comme mentionné plus haut, il est composé de brochures et pages webs visant à guider les justiciables dans des procédures judiciaires ou administratives, et de résumés de décisions de la CSC, à visée informative, rédigés par des experts du droit et destinés aux non-experts.

5.3.2 Le sous-corpus EN-CAB

Le sous-corpus Citizens Advice est composé de 515 textes relativement courts (997 mots en moyenne, formant un total de 711 935 mots (*tokens*), extraits du site web anglophone d'une association d'aide à l'accès au droit présente dans tout le

Royaume-Uni : le Citizens Advice Bureau. Ce corpus de textes, principalement procéduraux, porte sur différents sous-domaines du droit, et a pour but de guider les usagers du droits non-experts dans leurs démarches, qu’elles soient judiciaires et administratives. On trouve des textes procéduraux portant sur des sujets et démarches aussi divers que la protection du consommateur, le divorce et l’adoption, les procédures judiciaires. Ce corpus a été conçu car l’association Citizens Advice Bureau est particulièrement implantée dans les communes britanniques, et se réclame du style clair, comme l’affirme sa page d’accueil :

We always write in plain English and our tone of voice is straightforward, reassuring and positive. To make sure our writing is consistent we use a style guide. (Citizens Advice 2016)

Comme nous l’avons vu plus haut, le droit est intrinsèquement lié à un territoire. Il convient donc d’abord de souligner que le contenu du site web diffère selon que la personne le consultant se trouve en Écosse, au Pays de Galles, en Angleterre ou en Irlande du Nord, comme l’indique l’encart demandant de choisir, à l’arrivée sur le site, à quel système national de droit on se réfère. Ces différences juridiques et parfois linguistiques (le site possédant une version en gallois) sont dues au caractère quasi-fédéral du système législatif et judiciaire britannique, qui peut entraîner des différences dans la législation. Par mesure de simplicité, tous les textes compilés pour cette étude concernent l’Angleterre. Par ailleurs, comme discuté plus haut, la question de la temporalité entre en jeu, à plus forte raison dans un corpus de dissémination du droit expliquant la loi aux justiciables : les textes sont susceptibles d’être sujet à des changements de législation, la loi étant constamment modifiée, amendée, réécrite, abrogée, ou de nouvelles lois créées. Le sous-corpus CAB illustre bien cette dimension : composé de textes récoltés de décembre 2017 à octobre 2018, il reflète la diffusion du droit telle qu’il était sur cette période. De nouveaux textes sont venus s’ajouter à ceux déjà présents, reflétant les changements politiques et législatifs, notamment la de l’Union Européenne du Royaume-Uni. Enfin, il faut noter que si les textes du Citizens Advice Bureau se fondent sur des textes spécialisés appartenant au discours législatif, ils n’y renvoient qu’en de très rares occasions. S’ils s’appuient sur un discours primaire spécialisé, ils ne constituent pas la version simplifiée des textes de loi, contrairement aux Brochures présentées ci-dessous, mais s’appuient bien sur des discours spécialisés d’origine, le discours législatif et le discours administratif, qu’ils visent à rendre accessible aux usagers.

5.3.3 Le sous-corpus Brochures

Le sous-corpus de Brochures, est composé de textes issus d'institutions différentes et de pays différents. Pour la partie anglophone, Ces textes sont souvent désignés par les institutions par les termes *Leaflet*, *Handbook* ou *Legal Guide*. Ils ont été publiés par les autorités locales du Royaume-Uni, d'autres par la Nouvelle-Zélande, entre 2014 et 2018. Nous avons sélectionné, pour le versant anglophone de ce sous-corpus, plusieurs textes estampillés "style clair", ou issus d'institution ayant reçu cette distinction de la part d'organismes qui promeuvent le style clair. L'association Plain English Campaign au Royaume-Uni, ou WriteMark en Nouvelle-Zélande, proposent toutes deux d'évaluer le contenu produit par des institutions, compagnies et entreprises juridiques (d'assurances, notamment) s'adressant au grand public. Nous avons donc sélectionné des textes ayant reçu un tampon, souvent qualifiée de "marque", de la part de ces organismes présents dans différents pays anglophones. Les organisations promouvant le style clair s'associent en effet souvent à des institutions publiques, soit en leur offrant des formations, en les épaulant lors de la réécriture de textes et documents jugés complexes, ou, comme nous venons de le voir, pour les aider à évaluer le contenu de ces documents. Soit c'est le texte lui-même qui a reçu la marque ou le symbole de l'organisme qui évalue le style clair. C'est le cas notamment de textes ayant gagné les ClearMark Awards ou ClearWrite Marks. Certains des textes mentionnés au Chapitre 1 en font partie. Soit c'est tout le site web d'une institution juridique et administrative qui a reçu une récompense pour ses efforts de clarté, de la part des associations comme Plain English Campaign. Le site web de la municipalité de Belfast (Belfast City Council) déclare par exemple :

Plain English Campaign reviews our website each year to make sure its content, layout, navigation, language and design are easy to use and understand. This allows us to display the Internet Crystal Mark. (2019)

Parfois, il s'agit également de textes de médiation juridique d'institutions juridiques et administratives centrales, comme la NHS, qui nous ont été suggérés par des usagers du droit anglophone. Le sous-corpus anglophone, EN-Brochures, est donc composé de 45 textes publiés ou mis à jour en 2018, pour un total de 219 332 mots. Il porte sur des thèmes qui incluent la demande d'asile, l'adoption, l'assistance familiale, le divorce, mais également les droits des personnes arrêtées. Les textes, principalement procéduraux, ont été publiés par les institutions judiciaires (Ministère de la Justice ou équivalent) et administratives (autorités locales, régionales ou départementales). Ils font explicitement référence aux textes de lois qu'ils expliquent, et ont donc servi à construire le corpus législatif (voir plus bas).

La version francophone de ce corpus (ci-après FR-Brochures) est constituée de 27 textes (213 569 mots) publiés par les instances judiciaires et administratives françaises, publiés entre 2012 et 2018. Ils portent sur des procédures judiciaires semblables à celles de leurs homologues anglophones, bien que l'équilibre thématique ne soit pas total. Ces textes ont été publiés par les institutions, comme par exemple le Ministère de la Justice, mais aussi des institutions départementales ou municipales (comme la Ville de Paris), entre 2013 et 2018.

En moyenne plus longs que les textes du sous-corpus CAB, faisant entre 2 et 20 pages, ces textes ont donc été rassemblés sous l'étiquette "Brochures". Portant sur des procédures légales ou judiciaires comme l'adoption, ou informant les usagers de leurs droits, notamment en tant que témoin, victime ou détenu, ils réalisent à la fois une fonction informative et une fonction procédurale, et font explicitement référence aux textes de lois qu'ils expliquent. Ces brochures constituent donc la version simplifiée du discours législatif sur lesquelles elles s'appuient.

5.3.4 Le sous-corpus CA-Résumés

Ce sous-corpus est composé de résumés de décisions de justice de la Cour Suprême du Canada, qui ont à l'inverse des deux genres précédents, une fonction informationnelle. Il s'agit de résumés des décisions portant sur les affaires sur lesquelles la Cour a statué en 2018. Ils sont également un exemple de textes rédigés en style clair, comme le professe la page Web qui les présente vue au Chapitre 1.

Si la référence au langage clair est simple n'est pas explicite, nous avons choisi de compiler un corpus de ces textes car ils constituent la forme simplifiée d'un genre du discours juridictionnel, les décisions de justice, et sont issus d'une volonté de permettre aux usagers non-experts d'accéder à l'information juridique. Ce type de texte a attiré notre attention car il a été présenté à la conférence PLAIN de Montréal (2018) portant sur le PL en droit ; il n'a, à notre connaissance, jamais été étudié. Ces résumés sont également intéressants car ils font l'objet d'une traduction systématique vers le français. La partie anglophone de ce sous-corpus, EN-Résumés, est composée de 66 textes contenant 50 234 mots ; la partie francophone, FR-Résumés, contient quant à elle 62 182 mots dans 66 textes traduits. Ce sous-corpus est nettement plus petit que le précédent, ce qui est dû au genre discursif lui-même, les résumés contenant environ 942 mots (*tokens*) en moyenne, contre 4745 mots en moyenne par texte pour les textes du sous-corpus EN-Brochures (anglophone).

5.3.5 Les sous-corpus FR-DILA

Le sous-corpus FR-DILA est le pendant français du corpus EN-CAB, puisqu'il s'agit du site web de référence pour les usagers du droit francophone. Il est composé de sections portant sur différentes procédures judiciaires et administratives courantes auxquelles peuvent être confrontés les citoyen.nes résidant en France. Cependant, il n'a pas la même source discursive que ce sous-corpus anglophone. En effet, en France, c'est un acteur institutionnel qui prend en charge la diffusion du droit, la Direction de l'Information Légale et Administrative (DILA, d'où le nom du corpus), une institution qui dépend du cabinet du Premier ministre et publie des textes explicatifs du droit sur le site *service-public.fr*. Les textes composant ce sous-corpus sont à la fois des textes procéduraux et des textes explicatifs, rassemblés sous le titre "fiches d'information", expliquant aux citoyens soit les structures juridiques et administratives (par exemple les acteurs du domaine judiciaire), soit les guidant dans des procédures pour faire valoir leurs droits. Ce sous-corpus est composé de 712 textes, pour un total de 751 020 mots.

On notera les différences thématiques entre ces différents sous-corpus, issus de contextes juridiques et culturels divers. Par exemple, dans la catégorie "*Consumer*" du site CitizensAdvice, on trouve un texte procédural expliquant comment acheter une voiture d'occasion, une pratique réglementée. Suivant les conclusions de Pic et Furmaniak (2010), qui étudient la diffusion de textes spécialisés présentant un déséquilibre thématique similaire, cette asymétrie n'a pas été considérée comme un problème car elle reflète des différences dans la communication des institutions et organismes avec les usagers du droit.

5.3.6 Le Corpus LEX

Le corpus LEX est composé de quatre sous-corpus de textes juridiques présumés complexes, que nous comparerons aux textes du corpus PLAIN afin de caractériser le degré de simplification du droit dans le passage d'un discours normatif à un discours informatif destiné à un lectorat non-expert.

5.3.7 Les sous-corpus Lois

Ce sous-corpus est composé de textes législatifs anglophones et francophones. Dans notre cas, nous nous sommes concentrés sur le droit national, et avons donc ex-

clu les constitutions et traités internationaux, notamment les directives européennes. Les textes de lois choisis ne relèvent pas du droit des affaires mais davantage des sous-domaines du droit liés au droit public et privé, comme des questions pénales, judiciaires et administratives, incluant notamment le droit des étrangers et le droit familial, mais également la lutte contre les discriminations et l'attribution d'allocations. Le sous-corpus anglophone EN-Loi contient des textes issus des appareils législatifs néo-zélandais et britannique, en particulier ceux qui sont référencés dans les textes du sous-corpus EN-Brochures. Il compte seulement 35 textes pour près de deux millions de mots. Les textes de lois ou, *statutes*, sont en effet des documents de plusieurs centaines de pages, contenant non seulement des articles édictant la loi en question, mais également les amendements et modifications que cette loi entraîne sur les autres lois du corpus législatif. Les 35 textes du corpus EN-Lois comptent donc en moyenne 54 000 mots (*tokens*).

Concernant le versant en français, il comprend non des textes de lois entiers mais les articles qui sont systématiquement référencés dans les textes explicatifs du sous-corpus DILA. Les textes du corpus DILA font en effet référence à un ou plusieurs articles des différents Codes qui contiennent la loi. Le corpus FR-Lois contient donc 3 628 articles de loi, parfois très courts, pour un total de 4 551 673 mots. Contrairement aux textes de lois anglophones, collectés à la main, ce corpus a été compilé automatiquement à partir des hyperliens du corpus FR-DILA d'où la différence de taille entre les deux sous-corpus.

Comme mentionné plus haut, les textes du corpus LEX, en français comme en anglais, contiennent des articles et chapitres de loi qui datent de plusieurs décennies, mais qui étaient toujours en vigueur au moment de la constitution du corpus et référencés dans les textes du corpus PLAIN. C'est en effet ce dernier qui a été collecté en premier, et les textes du corpus LEX ont été sélectionnés car ils étaient cités comme référence dans certains textes du corpus EN-CAB et EN-Brochures.

5.3.8 Le sous-corpus CA-Jugements

Nous utilisons pour désigner ces textes le terme de *jugement*, utilisé par la CSC elle-même. A noter qu'en droit français, le terme *jugement* est réservé aux décisions des tribunaux, tandis que les décisions prises par les cours d'appel et de Cassation sont appelées *arrêts*. Ce corpus de jugements de la Cour Suprême du Canada, à mettre en regard du corpus CA-Résumés, est composé de deux sous-corpus :

- EN-CA-Jugements, composé de 49 jugements anglophones de la Cour Suprême du Canada correspondant aux affaires jugées en 2018
- FR-CA-Jugements, constitué des 46 traductions officielles en français de ces mêmes jugements.

Ces textes représentent donc à la fois un genre issu du discours juridictionnel, puisqu'ils sont produits par la plus haute juridiction du Canada, mais également des genres issus du discours judiciaire, puisque les affaires présentées devant la Cour ont d'abord été jugées par des institutions dites judiciaires (tribunaux judiciaires, tribunaux correctionnels, cours d'assises, tribunal de commerce...). Ces textes font également partie du discours normatif (au même titre que les lois) puisque comme nous l'avons vu dans le système de *common law* s'applique le principe du *Stare Decisis* : les décisions de la Cour Suprême font précédent, c'est-à-dire deviennent contraignantes pour les tribunaux inférieurs lors de jugement d'affaires similaires (tribunaux judiciaires et cours d'appels provinciales).

5.4 Résumé et outils de préparation des corpus

Dans cette section, nous présentons les statistiques descriptives des corpus LEX et PLAIN ainsi que les outils utilisés pour collecter et préparer les corpus avant leur analyse.

5.4.1 Statistiques

Dans les tableaux suivants, nous résumons les informations sur les données textuelles de nos corpus, incluant le nom du sous-corpus, le genre qu'il représente, ainsi que le nombre de mots (*tokens*) qu'il contient. Le premier tableau décrit le corpus PLAIN. Le second tableau concerne les textes du corpus LEX.

Bien que tous les textes du corpus PLAIN s'appuient sur un premier discours spécialisé réputé complexe (le genre juridique du *statute* ou de la décision de justice), seuls certains s'appuient explicitement sur un corpus de textes dont ils sont la version "traduites" en style clair. C'est le cas des sous-corpus EN-DILA et EN/FR-Résumés et EN/FR-Brochures. Le sous-corpus CAB, quant à lui, s'appuie implicitement sur des textes de loi (*statutes*); les textes constituent bien cependant des textes de diffusion du droit, et se réclament du *plain language*. Les corpus LEX et PLAIN, au bout du compte, sont constitués respectivement de 7 862 948 et 2 134 387 mots. La

TABLEAU 5.1 – Le corpus PLAIN

Nom du sous-corpus	Genre	Langue	Dates	Nombre de textes	Nombre de mots
EN-CAB	Textes explicatifs-procéduraux	Anglais	2018	773	711 935
EN-Brochures	Textes explicatifs-procéduraux	Anglais	2014-2018	45	219 332
EN-CA-Résumés	Résumés-informatifs	Anglais	2017-2018	66	50 234
Total EN-PLAIN				884	981 501
FR-DILA	Textes explicatifs-procéduraux	Français	2018	845	877 135
FR-Brochures	Textes explicatifs-procéduraux	Français	2013-2018	27	213 569
FR-CA-Résumés	Résumés-informatifs	Français	2017-2018	66	62 182
Total FR-PLAIN				938	1 152 886

TABLEAU 5.2 – Le corpus LEX

Nom du corpus	Genre	Langue	Nombre de textes	Nombre de mots	Dates
EN-Lois	Textes législatifs	Anglais	36	1 975 302	1986-2018
EN-CA-Jugements	Jugements	Anglais	49	671 124	2018
Total EN-LEX				85	1 152 886
FR-Lois	Articles de lois	Français	3628	4 506 108	1967-2018
FR-CA-Jugements	Jugements	Français	46	710 414	2018
Total FR-LEX				3 674	5 216 522

constitution de "super-corpus" de plusieurs millions voire milliards de mots semble caractéristique de la recherche en linguistique de corpus à l'ère du *Big Data* (Nelson (2010)). Cependant, ici, s'agissant de corpus spécifiques ou spécialisés, nous avons préféré nous pencher sur plusieurs corpus, de petite à moyenne taille et composé de peu de textes. (Koester 2010 : 70) rappelle en effet qu'un petit corpus spécialisé possède un avantage pour étudier les caractéristiques d'un registre ou d'un genre, puisque les schémas linguistiques observés peuvent ainsi être plus aisément replacés dans leur contexte d'utilisation et faire l'objet d'une analyse qualitative, que dans un corpus composé de plusieurs milliers de textes et de plusieurs milliards de mots. Les observations linguistiques sur un corpus représentant des usages spécifiques de la langue dans un contexte particulier sont par ailleurs stables dès que l'on obtient un échantillon de dix textes (Koester 2010 : 70), c'est pourquoi certains sous-corpus sont composés seulement de quelques dizaines de textes.

En termes de longueur des textes, il faut noter qu'il y a une certaine variabilité du nombre de mots par texte dans certains sous-corpus.

5.4.2 Outils de de saisie et préparation des corpus

Plusieurs méthodes ont été utilisées pour extraire les données textuelles, qui nous ont permis de récolter un grand nombre de textes. Le sous-corpus Brochures a été compilé à partir de recherches par mots-clés sur des sites institutionnels. Le plus souvent en pdf, ces textes ont donc été convertis en fichiers texte afin de pouvoir être pris en charge par des concordanciers et outils de textométrie. En revanche, des outils d'indexation et d'"aspiration" automatisées (*web crawling* et *web scraping*) ont été utilisés (sous RStudio) pour collecter les textes procéduraux du Citizens Advice Bureau et du Service Public français (FR-DILA), ainsi que les jugements et résumés de décisions des Cours Suprêmes du Canada. Les textes de FR-Lois ont été compilés et aspirés automatiquement à partir des liens présents dans les pages de FR-DILA, chacun de ces textes destinés au public faisant systématiquement référence aux articles de lois sur lesquels ils s'appuient. Concernant les textes de EN-Lois, ils ont été téléchargés manuellement en s'appuyant sur les références à des textes législatifs dans EN-Brochures et EN-CAB. Or, ce dernier, nous l'avons déjà évoqué, contient peut de référence aux lois sur lesquelles il s'appuie, bien qu'il s'agisse (comme FR-DILA) d'un corpus de médiation juridique simplifiant le discours du droit. Cette différence dans la collecte des données explique le déséquilibre entre le nombre de textes et le nombre de mots entre EN-Lois et FR-Lois.

Les textes du corpus, comprenant différents formats, ont tous été convertis en texte brut afin de pouvoir les traiter à l'aide des outils de la linguistique de corpus. Un nettoyage des textes a également été nécessaire, puisque les textes contenaient des mots, abréviations et segments de phrase susceptibles de représenter du "bruit", notamment lors de la création des listes de mots-clés ou de segments répétés (ngrams, collocations). Il est intéressant de noter que les segments à nettoyer étaient eux-mêmes liés à la structure et des différents genres discursifs. Ainsi dans les textes de diffusion du Citizens Advice Bureau ou du DILA, ce sont des termes ou expressions liées à la généricité du Web qui constituaient des éléments susceptibles de "bruit" les résultats statistiques, notamment les boutons cliquables permettant d'accéder à un formulaire administratif, mais surtout des liens hypertextes renvoyant vers d'autres textes explicatifs sur une procédure juridique connexe ou vers le site d'une autre institution. Concernant les textes du corpus LEX, ce sont des éléments liés à la nature profondément intertextuelle des discours législatif et judiciaire, qui réfèrent eux-mêmes au corpus législatif et jurisprudentiel. En particulier, les textes législatifs EN-Lois a dû être nettoyé, les textes de lois contenant une section décrivant les amendements qui produisent un effet sur d'autres lois ou régulations déjà intégrée

au corps législatif, auxquelles on renvoie sous la forme de nombreuses abréviations et codes. Dans notre cas, les textes contenaient un nombre d'occurrences importants de l'abréviation "S.I." pour *Statutory Instrument*, suivie de numéros, des codes qui renvoyaient aux régulations gouvernementales sur lesquelles la loi en question aurait prochainement effet. Les décisions de la Cour Suprême du Canada, quant à eux, ont dû être débarrassés des 3 336 occurrences de l'abréviation "v.", correspondant aux titres d'autres décisions citées dans le jugement, dont la forme est fixe : "Nom de l'Appelant (*Appellant v.* (en anglais)/ *c.* (en français) Nom de l'Intimé.e (*Respondent*)". Ces titres contiennent, lorsque c'est la Couronne qui est à l'origine de la présentation du cas devant la Cour Suprême, l'abréviation "R.", faisant référence à la Reine, qui a également été supprimée dans nos corpus de décisions, en français et en anglais. Nous avons donc retiré ces occurrences, afin notamment qu'elles ne puissent fausser les mesures se fondant sur les phrases et le nombre de mots par phrases, celles-ci étant repérées grâce aux points. Le nettoyage a été réalisé grâce aux expressions régulières, qui ont permis de sélectionner rapidement et précisément les chaînes de caractères à retirer.

Nous avons ensuite pu procéder à l'analyse du corpus grâce aux différentes approches et outils méthodologiques présentées au Chapitre 4. Nous présentons l'analyse de ces résultats dans la partie suivante.

Troisième partie

Analyse et discussion des résultats

Chapitre 6

Plain language et machine-learning : Exploration d'un modèle d'apprentissage supervisé pour caractériser le discours juridique avant et après simplification

Dans ce chapitre, nous présentons les résultats de la classification automatique par un modèle, qui vise à assigner une étiquette aux textes, LEX ou PLAIN, à partir des métriques calculées sur ces textes. Pour qu'une classification soit considérée réussie, le modèle devra avoir attribué au texte LEX ou PLAIN une étiquette correspondant à son corpus d'appartenance. L'objectif est d'évaluer le caractère prédictif des métriques de complexité comme *features* dans le cadre d'une approche d'apprentissage supervisé. Une analyse des résultats permet ensuite d'analyser quels types de métriques sont les plus performantes pour différencier les textes spécialisés de ceux destinés aux justiciables, et ce qu'elles révèlent sur les caractéristiques des deux corpus. Dans la première section, nous présentons les résultats de la classification et une description quantitative des métriques sélectionnées par le modèle. Nous nous tournons ensuite vers une analyse qualitative des scores de lisibilité et de la longueur moyenne des unités syntaxiques (Section 2) puis des métriques morpho-syntaxiques et lexicales perçues comme prédictives (Sections 3), afin d'examiner ce

que ces métriques peuvent révéler sur le passage du discours juridique au discours de reformulation du droit. Dans la Section 4, en nous fondant sur les limites méthodologiques, nous discutons des apports et des problèmes posés par notre approche quantitative et *machine-learning* du style clair.

6.1 Résultats quantitatifs

Comme nous l’avons vu au Chapitre 5, le modèle d’apprentissage supervisé classe les textes en se fondant sur les mesures. Pour cela, il divise le jeu de données en deux jeux d’entraînement (c’est-à-dire avec les étiquettes LEX ou PLAIN) et un jeu de test (sans les étiquettes). Les résultats sont donnés sous la forme d’une matrice de confusion, c’est-à-dire qui présente le nombre de textes dont la classe a été correctement prédite, et ceux qui n’ont pas été bien classés, avec des scores de performance que nous présentons.

6.1.1 Performance de la classification des textes LEX et PLAIN à partir des métriques de complexité

Les résultats du Tableau 6.1 montrent les scores moyen pour l’*accuracy*, le rappel et la précision sur les données d’entraînement et de test *training/test sets*. L’*accuracy*

TABLEAU 6.1 – Performance de la classification automatique des textes en se fondant sur les métriques de complexité (en %)

Langue des textes	Accuracy	Précision	Rappel
anglais	97	94	75
français	73	91	54

représente le pourcentage de bonnes prédictions. Elle est de 97% pour le jeu de donnée en anglais, ce qui signifie que le modèle fait presque un sans-faute : sur 948 textes pour les deux corpus EN-PLAIN et EN-LEX, 20 ont été mal classés. Pour le français, l’*accuracy* est de 73%. Cela signifie que 73% des textes sont bien classés : sur les 884 textes du jeu de données, 200 ont été mal classés. L’algorithme apparait donc meilleur pour prédire la classe des textes rédigés en anglais. Les deux autres scores

permettent de déterminer plus précisément quels types d'erreurs le modèle effectue. La précision (dans la deuxième colonne du Tableau 6.1) correspond au pourcentage de textes appartenant vraiment à la catégorie LEX parmi tout les textes ayant été prédits comme LEX par le modèle. Il représente donc la proportion de textes appartenant à la catégorie LEX sur la somme des textes "véritablement LEX" et des textes PLAIN prédits comme LEX, ces derniers étant communément appelés faux positifs. En comparant ces scores pour les deux langues, qui sont respectivement de 94% et de 91%, on constate que l'algorithme est à peu près aussi précis pour l'anglais que pour le français : dans les deux cas, peu de textes PLAIN ont été prédits comme LEX. En revanche, on note une nette différence dans le rappel (*recall*), entre les deux langues. Le rappel est également appelé sensibilité ; il s'agit du pourcentage de textes prédits comme LEX parmi tous ceux qui appartiennent vraiment à la catégorie LEX. La sensibilité est de 75% pour le jeu de données anglophones, ce qui montre que peu de textes de la catégorie LEX sont mal classés. L'algorithme arrive donc (relativement) bien à différencier les deux classes en anglais. En français, la sensibilité est au contraire de 54%, ce qui signifie que l'algorithme peine à trouver une partie des textes LEX (qu'il classe PLAIN). Nous discutons des raisons possibles pour cette différence entre les deux performance dans la Section 4.

La première conclusion est que les métriques, en particulier pour l'anglais, ont un bon pouvoir de prédiction de la classe des textes : le modèle parvient globalement assez bien à assigner le bon corpus (LEX ou PLAIN) au texte en se fondant sur ces métriques. Nous pouvons maintenant nous tourner vers ce qui nous intéresse le plus dans cette classification : quelles ont été les métriques les plus pertinentes pour classer les textes dans les catégories LEX ou PLAIN ? Que permettent-elles de dire sur le passage du style juridique au style clair ?

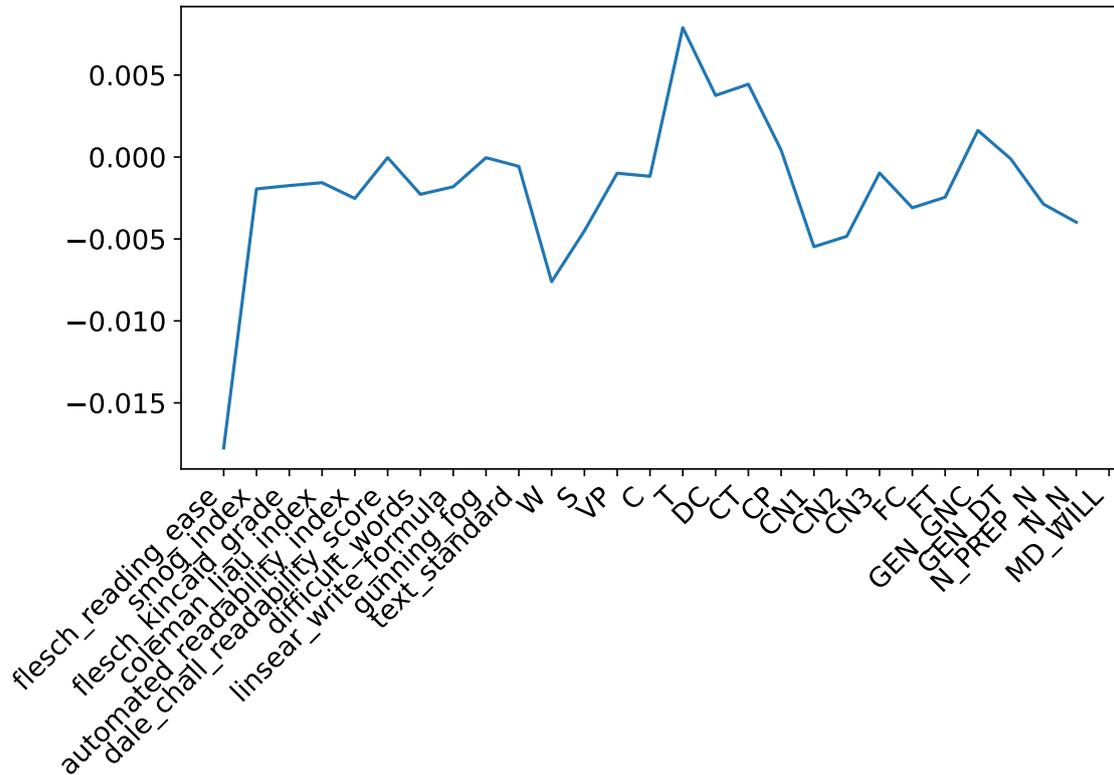
6.1.2 Quelles sont les métriques de complexité les plus puissantes pour prédire le corpus d'origine d'un texte ?

La régression avec régularisation permet, nous le rappelons, de sélectionner les *features* les plus importants pour la classification, c'est-à-dire qui donnent le plus d'information à l'algorithme pour prédire la classe d'un texte en se fondant sur lesdites métriques. Le choix du paramètre C permet ensuite d'obtenir un petit nombre de métriques qui ont le plus grand pouvoir de prédiction. Ici, le paramètre choisi est $C = 0,001$, qui a permis de déterminer une dizaine de métriques ayant un fort pouvoir de prédiction.

Métriques ayant le plus fort pouvoir de prédiction pour les corpus EN-LEX et EN-PLAIN

Les poids des métriques pour la classification des textes anglophones sont présentées dans la Figure 6.1. Les métriques ayant un coefficient positif (vers le haut

FIGURE 6.1 – Poids des *features* dans la classification des corpus EN-LEX et EN-PLAIN



sur la Figure 6.1) sont associées à la prédiction de la classe LEX, c'est-à-dire qu'ils sont perçus par l'algorithme comme caractéristiques des textes du corpus EN-LEX. Les métriques apparaissant avec un coefficient négatif sur la figure sont quant à elles prédictives pour la classe PLAIN. Les métriques ayant obtenu un score de 0 (donc ayant un faible pouvoir de prédiction) ne sont pas représentées sur la figure.

Pour plus de lisibilité, nous reproduisons les dix métriques les plus prédictives dans les Tableaux 6.2 et Tableau 6.3, en indiquant pour chacune d'elles l'outil qui les a mesurées dans la chaîne de traitement, quelle dimension de la complexité cet outil capture, et pour quelle catégorie (LEX ou PLAIN) cette métrique est prédictive.

Les résultats suggèrent tout d'abord que pour l'anglais, la plupart des métriques les plus importantes sont caractéristiques des textes du corpus PLAIN, puisque

TABLEAU 6.2 – Métriques les plus prédictives pour la classe EN-PLAIN

Métrique	Outil de mesure	Dimension mesurée
Flesch Reading Ease	Texstat	Longueur moy. phrases et mots
Verb Phrases (GV)	L2SCA	Nombre de groupes verbaux
Mean Length of Sentence (MLS)	L2SCA	Longueur moy. unité syntaxique
Mean Length of T-Unit (MLT)	L2SCA	Longueur moy. unité syntaxique
Clause	L2SCA	Nombre de propositions

davantage de métrique ont un coefficient négatif associé, comme le montrent la Figure 6.1. La métrique qui est particulièrement puissante pour prédire la classe PLAIN est le score de lisibilité Flesch-Reading Ease, dont nous avons parlé au Chapitre 3. Ce score s’appuie sur la longueur moyenne des mots (en syllabes) et la longueur moyenne des phrases ; il n’est donc pas surprenant que la métrique MSL (*Mean Length of Sentence*), c’est-à-dire la longueur moyenne des phrases, apparaisse comme la troisième la plus prédictive pour la classe PLAIN. Parmi les outils qui ont fourni les métriques pour prédire les textes EN-PLAIN, c’est l’analyseur de complexité syntaxique L2SCA qui est le plus représenté. En plus de la longueur moyenne des phrases, la longueur moyenne des T-Units, le nombre de propositions par phrase et le nombre de Groupes verbaux (Verb Phrases) apparaissent comme caractéristiques des textes PLAIN pour le modèle. Ce sont dans en particulier les métriques associées à la complexité syntaxique qui sont prédictives pour les textes EN-PLAIN. Concernant la classe LEX pour le corpus anglophone, les métriques associées par le modèle sont présentées dans le Tableau 6.3.

TABLEAU 6.3 – Métriques les plus puissantes pour prédire la classe EN-LEX

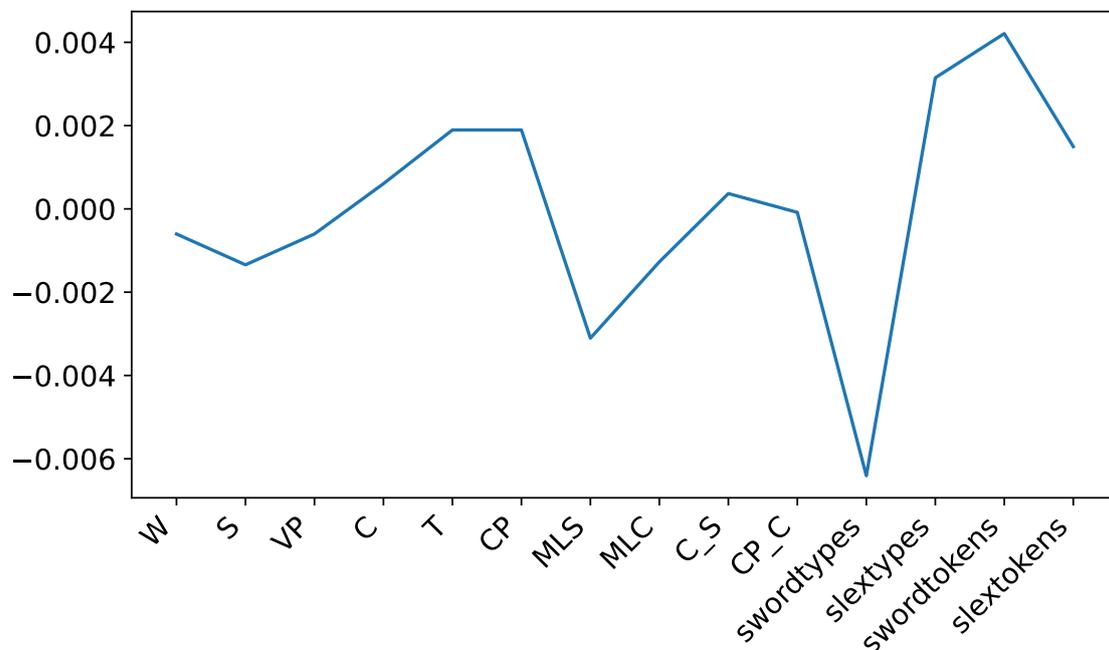
Métrique	Outil de mesure	Dimension mesurée
Complex Nominals 1	L2SCA	GN pré- et post-modifiés
Article THE	Microsystems	Proportion d’articles définis (Gaillat et al. 2022)
N_PREP_N	Microsystems	N FOR/OF N

Sur une dizaine des 143 variables utilisées pour les corpus anglophones, trois métriques apparaissent, selon le modèle, comme ayant un fort pouvoir de prédiction pour classer les textes dans cette catégorie. Il est intéressant de noter que deux d'entre elles font partie de la version modifiée du L2SCA, les métriques appelées microsystèmes, qui capturent certains ratios de formes par rapport à d'autres. Il s'agit des métriques qui capturent le nombre d'articles définis *Article THE* et de la métrique N_PREP_N, un score qui se fonde sur le nombre de groupes nominaux complexes composés de deux GP reliés par les préposition OF ou FOR. De plus, la métrique Complex Nominals, capturée par l'analyseur de complexité syntaxique L2SCA, qui se fonde sur la pré- ou post-modification des groupes nominaux, est également sélectionnée par le modèle comme aidant beaucoup pour classer les textes LEX.

Prédicteurs les plus performants pour les corpus francophones

La Figure 6.2 montre les *features* (métriques) ayant le plus fort pouvoir de prédiction d'après le modèle.

FIGURE 6.2 – Poids des *features* dans la classification des corpus FR-LEX et FR-PLAIN



Nous rappelons que pour le français, seules 49 métriques ont été mesurées,

puisque les métriques "Microsystèmes" n'ont pas été adaptées pour le français. Seuls les outils LCA (complexité lexicale), L2SCA (complexité syntaxique) et Textstat (scores de lisibilité) ont été utilisés ici.

TABLEAU 6.4 – Métriques ayant le plus fort pouvoir de prédiction pour les textes FR-PLAIN

Métrique	Outil de mesure	Dimension mesurée
Flesch Reading Ease	Textstat	Longueur moy. phrases et mots
Verb Phrases (GV)	L2SCA	Nombre de groupes verbaux
Sophisticated Word Types	L2SCA	Nombre de types de mots peu fréquents

D'après le modèle, comme présenté dans le Tableau 6.4, trois métriques sont associées aux textes FR-PLAIN : le score FRE, le nombre de groupes verbaux, comme pour l'anglais, mais également une métrique du LCA associée à la sophistication lexicale, en particulier aux mots n'apparaissant pas dans les 2 000 mots les plus fréquents (Lu 2012). On voit dans le Tableau 6.5 que le modèle associe davantage de métriques aux textes issus du corpus FR-LEX.

TABLEAU 6.5 – Métriques ayant le plus fort pouvoir de prédiction pour les textes FR-LEX

Métrique	Outil de mesure	Dimension mesurée
Automated Readability Index	Textstat	Longueur moy. phrases et mots
Flesch Kincaid Grade	Textstat	Longueur moy. phrases et mots
Mean Length of Sentence	L2SCA	Longueur moy. phrase
Mean Length of Clause (MLC)	L2SCA	Longueur moy. proposition
Linsear Write Formula	Textstat	Score fondé sur nb de syllabes / mot et mot / phrase
Sophisticated Lexical Types	L2SCA	Nb de types de mots lexicaux peu fréquents
<i>Coordinate Phrase</i> (CP)	L2SCA	Nb coordinations par proposition

Les métriques associées par le modèle sont principalement des scores de lisibilité : l'*Automated Readability Index*, le *Flesch-Kincaid Grade* et la *Linsear Write*

Formula indiquent tous un score correspondant au niveau scolaire du système éducatif américain nécessaire (selon le lectorat pris comme modèle, correspondant à la figure des lecteur.ices "moyen.nes") pour lire facilement le texte ; le premier se fonde sur le nombre de caractères par mot et de mots par phrase, le second renvoie un score fondé sur le nombre de mots par phrase et de syllabes par mot ; et le troisième sur la proportion de mots de deux syllabes ou moins ou de trois syllabes ou plus pour un échantillon de 100 mots. Trois des métriques sélectionnées par le modèle sont mesurées par l'analyseur de complexité syntaxique : la longueur moyenne des phrases, la longueur moyenne des propositions (*Mean Length of Clause*) et la proportion de coordination (*coordinate phrases*). Enfin, comme pour le corpus FR-PLAIN, le modèle sélectionne une métrique de complexité lexicale, *Sophisticated Lexical Types*, qui s'intéresse au nombre de mots lexicaux considérés comme sophistiqués.

6.1.3 Description statistique des métriques les plus prédictives

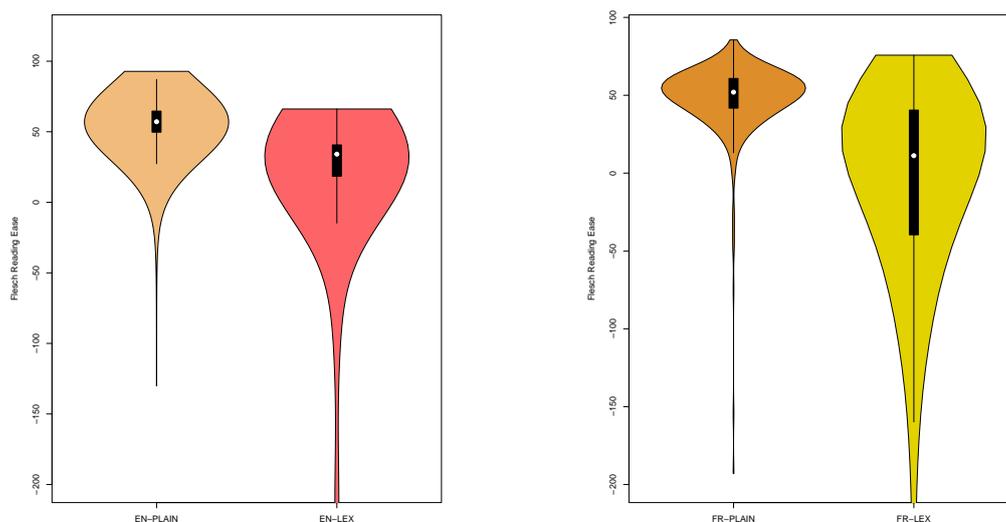
Nous décrivons d'abord quantitativement la distribution des scores obtenus par les textes pour les métriques présentées ci-dessus. Dans les sections suivantes, nous effectuons une analyse qualitative des métriques, exemples à l'appui.

La première métrique pointée par le modèle comme étant très prédictive pour distinguer les textes LEX de PLAIN, en français comme en anglais, est le score Flesch Reading Ease (FRE). Le Tableau 6.6 montre la répartition des scores pour chaque corpus.

TABLEAU 6.6 – Score Flesch Reading Ease dans les corpus LEX et PLAIN

Corpus	Médiane	Ecart-type	Minimum	Maximum
EN-PLAIN	56,6	15,3	-130,2	92
EN-LEX	34,1	96,9	-570,0	66,1
FR-LEX	14,1	50,4	-1449,4	75,4
FR-PLAIN	52,3	13,1	15,2	85,5

D'après les résultats, les scores FRE dans le corpus EN-PLAIN sont distribués autour de 56,6. Cela signifie que la moitié obtiennent un score de 57 ou au-dessus,



(a) Score de lisibilité FRE dans les corpus EN-PLAIN et EN-LEX (b) Score de lisibilité FRE dans les corpus FR-PLAIN et FR-LEX

FIGURE 6.3 – Score de Flesch Reading Ease dans les quatre corpus

soit environ 60, comme préconisé pour rédiger en PL. L'écart-type de 15,3 montre que les scores sont somme toute assez homogènes, tandis que ceux du corpus EN-LEX sont caractérisés par une très grande variabilité, avec un écart-type de 96 et un véritable grand écart entre les valeurs maximales et minimales. Certains textes des corpus EN-LEX ET FR-LEX obtiennent des scores négatifs extrêmement bas, avec un minimum de -570 et -150 respectivement. L'échelle a en effet une limite haute de 100, mais pas de limite basse, il est donc possible qu'un texte obtienne un score négatif de par la façon dont il est calculé, la formule incluant une soustraction (voir Chapitre 3, p.139 pour la formule détaillée). Le score de FRE est la métrique la plus fortement associée par le modèle d'apprentissage au corpus PLAIN, car, malgré des exceptions, il reste relativement stable et haut en moyenne, par comparaison avec le corpus EN-LEX. Pour ce dernier en effet, la médiane des scores obtenus est de 34,1, et est quant à lui caractérisé, nous l'avons évoqué, par une plus grande variabilité des scores de lisibilité. Environ 40% des textes du corpus EN-LEX obtiennent des scores FRE inférieur à 30. Cette nette différence dans la distribution des scores entre les deux corpus peut expliquer pourquoi cette métrique a été repérée par le modèle comme particulièrement discriminante.

Le score FRE est également associé aux textes du corpus FR-PLAIN. Dans le Tableau 6.6, on voit que les scores de Flesch Reading Ease dans le corpus FR-PLAIN sont ramassés autour de la valeur médiane de 52, avec quelques exceptions : neufs

textes du corpus FR-PLAIN obtiennent un score négatif, et un texte a même un score FRE de -130,2. Les scores FRE des textes du corpus FR-LEX sont distribués autour de la médiane de 11,9, et sont donc beaucoup plus bas. Là encore, l'écart-type de 50,4 indique que le jeu de données FR-LEX est plus hétérogène, avec des scores qui sont à la fois plus bas et plus variables, contrairement aux scores FRE pour le corpus FR-PLAIN, globalement plus haut et plus homogènes, ce qui explique pourquoi cette métrique a un fort pouvoir de prédiction pour les textes francophones FR-PLAIN.

Le Tableau 6.7 indique les scores médians et l'écart-type des métriques les plus importantes pour prédire la classe des textes anglophones d'après l'algorithme.

TABLEAU 6.7 – Médiane et écart-type des métriques de complexité lexicale et syntaxique les plus prédictives pour l'anglais

Corpus	Verb Phrases	Mean Length of Sentence	Mean Length of T-Unit	Clause	Complex Nominals	Article THE	Squared Verb variation	N_PREP_N
EN-LEX	56 (1031)	28 (11)	27 (11,2)	39 (733)	101 (1323)	31 moy 375 (569)	36,6 (28,5)	37 (390)
EN-PLAIN	106 (114)	22 (6,9)	20 (6,6)	75,5 (79,4)	64 (89)	21 (33)	20,5 (9,8)	11 (22)

Ce que l'on note d'abord, c'est que les valeurs pour ces scores sont beaucoup plus resserrés au niveau des valeurs médianes pour les textes EN-PLAIN que pour les textes EN-LEX, qui présentent, comme le score FRE, une dispersion très grande. Seules les métriques *Verb Phrases* et *Clauses* ont une médiane plus élevée pour le corpus EN-PLAIN. L'écart-type (présenté entre parenthèse dans le tableau) révèle une très grande hétérogénéité des scores pour les textes juridiques ; comme nous le verrons plus loin, cela peut être dû à des différences à l'intérieur du jeu de données lui-même. Ces différences fortes entre les corpus EN-PLAIN et EN-LEX peuvent expliquer que certaines métriques aient été détectées comme ayant un fort pouvoir de prédiction.

Le Tableau 6.8 donne les résultats de mesures statistiques descriptives menées sur les métriques syntaxiques et lexicales qui ont permis, d'après le modèle, de classer les textes du corpus francophone.

TABLEAU 6.8 – Médiane et écart-type des métriques de complexité lexicale et syntaxique les plus prédictives pour le français

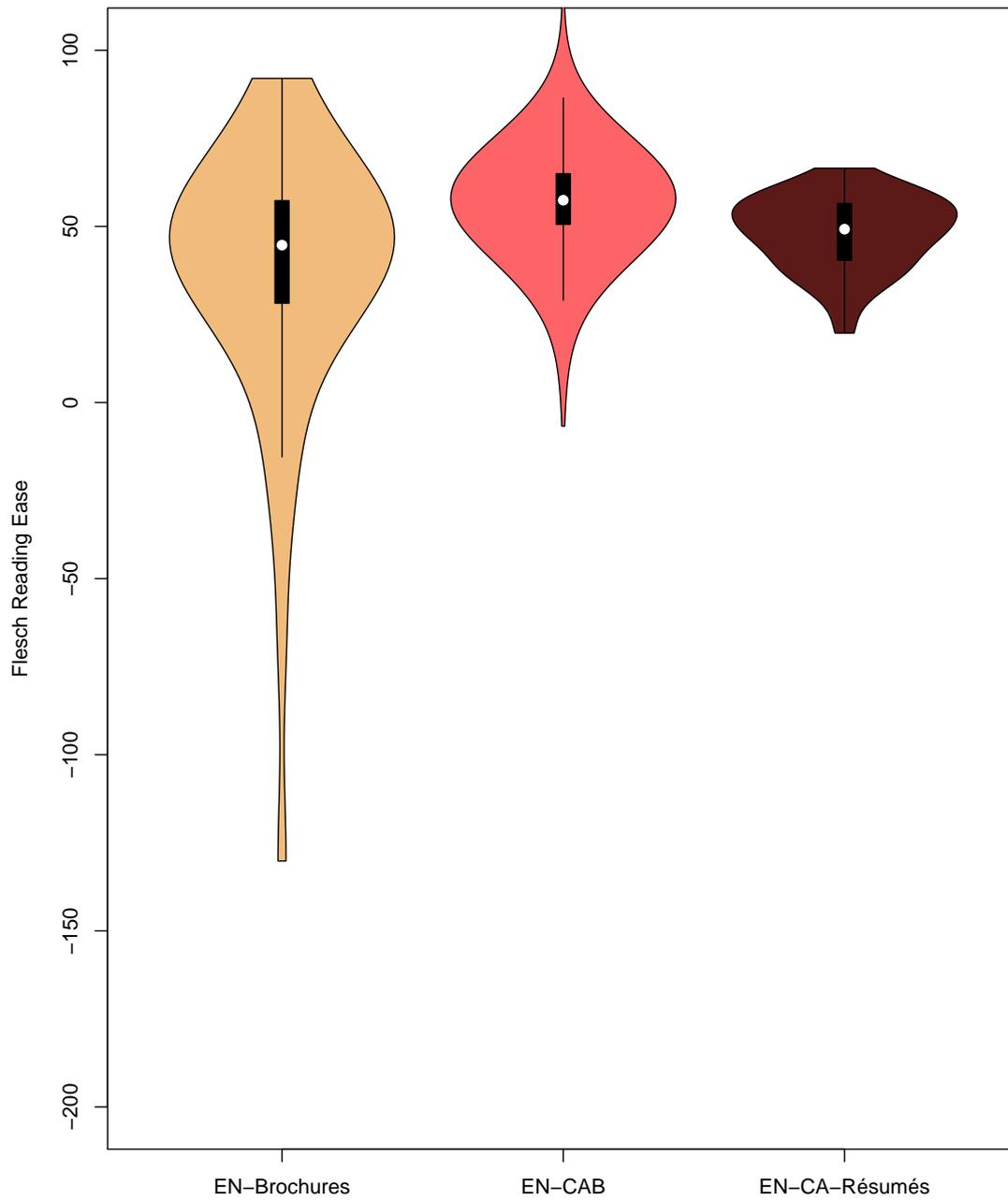
Corpus	Verb phrases	Sophisticated Word Types	Sophisticated Lexical Types	Coordinate Phrases
FR-LEX	12 (202)	127 (333)	78 (232)	51 (823)
FR-PLAIN	40 (62)	209 (157)	129 (116)	107 (219)

Ici, on voit que le score médian pour les métriques *Verb Phrases* est plus de trois fois plus haut pour FR-PLAIN que pour FR-LEX et que le score *Sophisticated Word Types* est presque deux fois plus haut pour ce corpus que pour FR-LEX. Là encore, cela peut expliquer que ces métriques aient été repérées comme ayant un fort pouvoir de prédiction pour la classe PLAIN en français. Comme pour l'anglais, on voit que la distribution des données (en particulier l'écart-type entre parenthèse) indique que le corpus FR-LEX présente une grande variation. Que suggèrent ces résultats ? Globalement, les tendances centrales représentées par la médiane, et la dispersion montrent que les textes du corpus PLAIN (dans les deux langues) sont plus homogènes que les textes du corpus LEX.

6.1.4 Y-a-t-il des différences entre les genres discursifs du corpus PLAIN ?

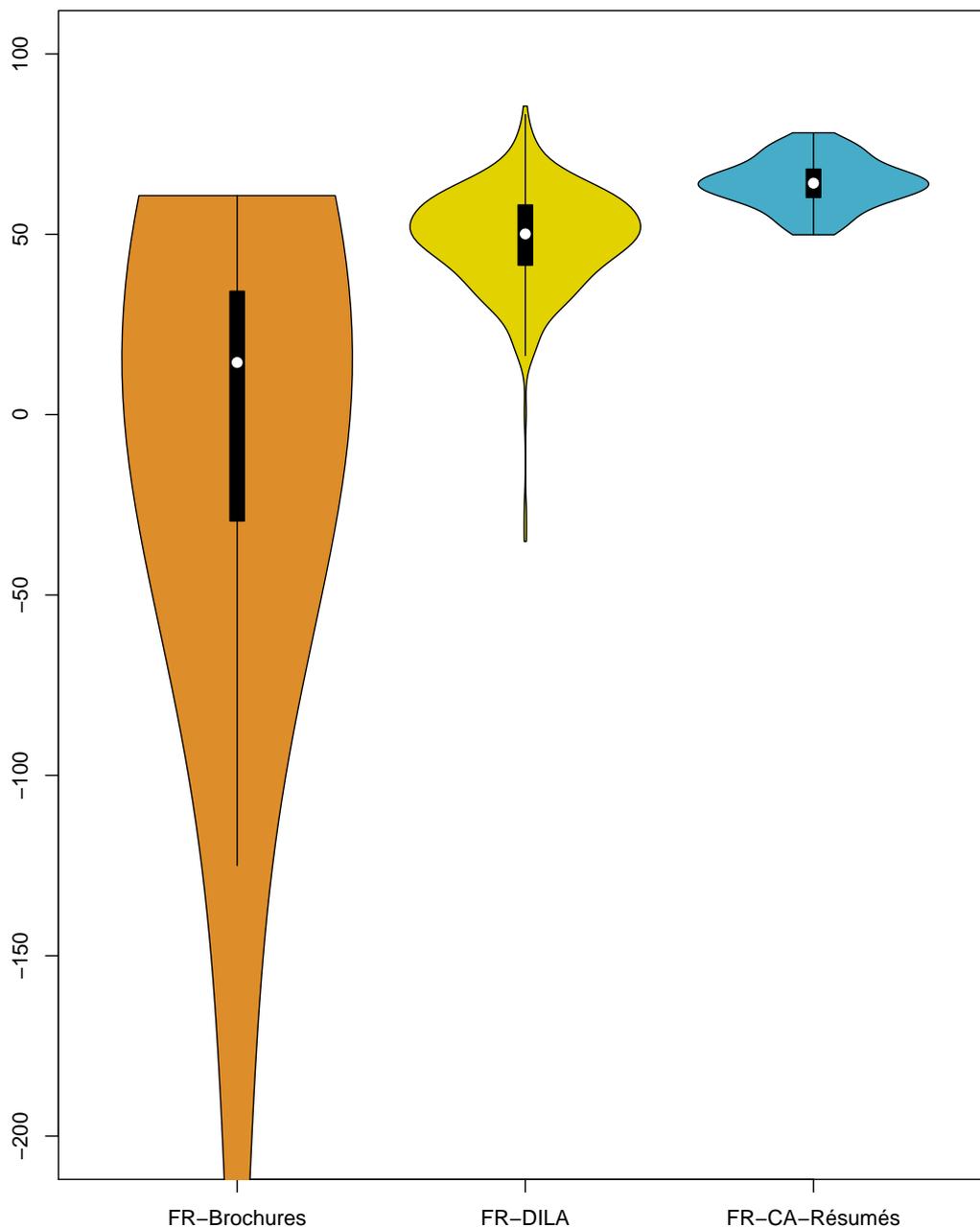
Les résultats présentés jusqu'ici portent sur les corpus LEX et PLAIN entiers, sans différencier les sous-corpus, puisque c'est ce jeu de données qui a été utilisé pour le modèle. Nous avons cependant souhaité regarder les différences à l'intérieur du corpus PLAIN pour la métrique considérée comme la plus prédictive, c'est-à-dire le score FRE. Pour l'anglais, des différences nettes apparaissent entre les différents sous-corpus. Plus précisément, la médiane des scores FRE du corpus EN-Brochures est de 44,7, celle du corpus EN-CAB est de 57,4 et enfin celle du corpus EN-Résumés est de 49,2. On voit surtout des différences dans les valeurs minimales et maximales. Si le score FRE reste entre pour le corpus EN-CA-Résumés, on voit que pour les sous-corpus EN-CAB et EN-Brochures, la variabilité est beaucoup plus importante. Pour EN-Brochures, le score descend à -130,6 ce qui évoque certains scores du corpus EN-LEX. Le score minimal pour EN-CAB est également négatif

FIGURE 6.4 – Scores FRE - Différences entre les genres discursifs du corpus EN-PLAIN



(-6,7), avec un score maximal de 92. La Figure 6.5 montre quant à elle la différence entre les genres discursifs du corpus FR-PLAIN.

FIGURE 6.5 – Scores FRE - Différences entre les genres discursifs du corpus FR-PLAIN



On constate le même résultat pour les sous-corpus francophones. Pour le sous-corpus FR-CA-Résumés, le score FRE est compris entre 49,9 et 78,1, et reste donc assez stable autour de la valeur médiane de 64,2. A l'inverse, le corpus FR-Brochures présente une grande variabilité, avec un texte obtenant un score négatif de -554,7. Le corpus FR-DILA obtient également un score négatif de -35,2, avec une médiane 50,1.

Ces résultats suggèrent que ces différences sont dues à une différence générique : les corpus EN-CAB et EN-Brochures font partie du même sous-registre des textes procéduraux, celui des textes qui visent à guider le justiciable dans les procédures administratives. Les sous-corpus CA-Résumés, quant à eux, sont composés de textes qui résument les motifs de jugements de la CS du Canada. Il est également possible que cela soit dû simplement à une différence entre la façon dont les corpus ont été compilés. Les sous-corpus EN-Brochures et FR-Brochures sont issus de différentes institutions, qui pourrait expliquer la variabilité des scores. Nous discutons plus en détail des hypothèses concernant cette possible différence entre les genres discursifs dans la section suivante.

Nous avons présenté les résultats d'un point de vue quantitatif. Il faut maintenant revenir aux textes eux-mêmes pour nuancer et tenter d'illustrer qualitativement certains de ces résultats quantitatifs . Dans la suite de ce chapitre , les exemples cités sont présentés selon le sous-corpus auquel ils appartiennent.

6.2 Flesch Reading Ease et longueur moyenne des unités syntaxiques, des prédicteurs puissants d'après le modèle

Cette section est consacrée au score de lisibilité *Flesch Reading Ease* et à la longueur moyenne des phrases (MSL) et des unités syntaxiques (propositions ou T-units) désignées comme les plus prédictives par le modèle d'apprentissage. Ces métriques, très classiques nous l'avons vu, sont présentées quantitativement, puis sont illustrées, dans un mouvement du quantitatif au linguistique, par des extraits de corpus. Dans les exemples de cette section, le cas échéant, nous utilisons le symbole // pour signaler une proposition subordonnée et [] pour une proposition enchâssée.

6.2.1 Le score Flesch Reading Ease, le plus puissant prédicteur pour les corpus LEX et PLAIN

Le score de Flesch Reading Ease apparaît comme l'une des caractéristiques *features* les plus puissantes pour prédire la classe des textes PLAIN. Il s'agit d'un score souvent utilisé, nous l'avons vu, pour déterminer des seuils de lisibilité à atteindre pour les institutions juridiques ou dans le cadre de la diffusion de connaissances spécialisées. Pour chaque langue, nous présentons des extraits de textes des corpus LEX et PLAIN, afin d'illustrer ces métriques.

Que signifient les résultats quantitatifs du score FRE en termes de lisibilité ? Pour donner un ordre d'idée, Flesch (1979) indique le score FRE d'extraits aléatoirement choisis de certaines célèbres publications et revues, mais aussi de magazines anglophones dont il a calculé le score. Nous l'avons vu, les résultats sur le corpus EN-PLAIN indiquent que la moitié des textes ont un score supérieur à 57 sur l'échelle, ce qui correspond à un niveau de lecture pour des élèves de lycée, tandis qu'un score compris entre 60 et 70 correspond au PL et au niveau de lecture d'une personne de 13 à 14 ans ayant suivi une scolarité classique. Ces textes correspondent bien au niveau recommandé. A l'inverse, les résultats révèlent que la moitié des textes obtiennent un score *inférieur* à 57. Ainsi, plus de deux cents textes obtiennent un score en-dessous de 50, correspondant, toujours selon l'échelle de Flesch, au niveau de lecture d'une personne allant à l'université. Nous présentons quelques extraits de textes pour illustrer ces différences à l'intérieur du corpus EN-PLAIN.

Un score de 82 correspond selon ces mesures à la lisibilité des publicités dans un magazine, tandis qu'un score de 92 équivaut à celui d'une bande dessinée, d'après les échantillons mesurés par Flesch (1979). Il est notable que le texte qui obtient le score FRE le plus haut dans le corpus EN-PLAIN, exactement ce score de 92, est une brochure publiée par le ministère de la Justice néo-zélandais et intitulé *Children's guide to separation*, dont nous reproduisons un extrait ici :

(47) How do we sort out where I'll live and other stuff?

Your parents will usually work out what they think is best for you. They'll do this after asking you what you want. If they don't ask, let them know if you want to have a say.

"I spend about a week with each parent. I keep half my things at each house. It was tricky at first but now it works." "I like having two homes." "Don't be afraid to tell them what you need. Tell them what you think will work for you." (EN-Brochures)

Il s'agit du seul texte du corpus explicitement destiné à un lectorat mineur, il ne paraît donc pas accidentel qu'il obtienne un score aussi haut, qui témoigne d'une plus grande lisibilité. Converti en *grade level*, un score de 92 correspond en effet à l'âge de

lecture d'une personne de 10 ans en cursus standard (CM2 dans le système éducatif français). Cet exemple illustre plusieurs points intéressants. Tout d'abord, on voit que le nombre de mots par phrase reste en-dessous de 14 mots dans cet extrait. Par ailleurs, si le score FRE est si haut, c'est aussi parce que le texte est caractérisé par l'utilisation de mots monosyllabiques (*do, this, what, best*) ou polysyllabiques (*after, asking*) qui sont à la fois très courts et très fréquents (*high frequency words*). L'extrait comprend également des traits caractéristiques de la diffusion du discours spécialisé : on note l'emploi des pronoms personnels de deuxième et première personne (*you, I*), et l'utilisation de citations d'autres enfants et de témoignages pour permettre au lectorat (mineur, donc) de comprendre une situation juridique ou administrative qui les touche et a un impact sur leur quotidien (ici la garde partagée ou alternée après la séparation des parents). Cet exemple, bien qu'extrême, illustre parfaitement l'utilisation du PL, à l'attention d'un lectorat bien précis. On peut le comparer à l'extrait ci-dessous, qui est tiré d'un texte du corpus EN-PLAIN qui a obtenu 34 sur l'échelle FRE.

- (48) Generally Citizens Advice Bureaux must keep anything you tell them confidential. However, in cases// where there are concerns about child abuse, // they may have // to report it to the police or the local authority. This is // because they must report cases // if you are in danger or // if someone else is in danger. (EN-CAB).

La différence avec l'exemple (47) est ici évidente, puisque l'exemple (48) comprend des phrases beaucoup plus longues, notamment constituées d'énumérations de GN complexes (*the police or the local authority*), ainsi que de plusieurs mots de plus de quatre syllabes comme *confidentiality, authority*. Ces caractéristiques peuvent expliquer pourquoi le score descend en-dessous de 40. On trouve également, nous l'avons dit, des textes avec un score négatif. Nous discutons de ces exemples plus loin, car ils ont pu entraîner des erreurs de classification du modèle. Vis-à-vis des recommandations pour un style clair, qui préconisent un score de 60 ou plus, le score médian indique que la moitié des textes du corpus EN-PLAIN ne correspondent pas à un niveau "clair et simple" selon l'échelle. La plupart des textes d'EN-PLAIN, d'après ce score, ne correspondent donc pas à la norme recommandée par certaines institutions et par la littérature sur la lisibilité. Cependant, ce résultat est peut-être dû à la manière dont le score est conceptualisé et calculé, comme nous en discutons plus loin. Pour le français, la moitié des textes ont un score supérieur à 52. En particulier, 117 textes sur 424 obtiennent un score supérieur à 60, c'est-à-dire correspondant au style clair. L'extrait ci-dessous est issu d'un texte qui a obtenu

un score de 72,4.

- (49) Au Canada, de nombreuses personnes sont gardées en prison en attendant leur procès. Chaque jour, environ la moitié des gens incarcérés au Canada attendent leur procès et ne sont pas déclarés coupables. Il est plus difficile pour eux de se défendre quand ils sont incarcérés. Les gens détenus avant leur procès sont plus susceptibles de plaider coupable. La détention nuit également à leur bien-être physique et mental ainsi qu'à leur vie de famille et à leurs perspectives d'emploi. (EN-CA-Résumés)

Cet extrait contient plusieurs mots de trois syllabes (*susceptibles, détention, perspective*) mais il obtient un score considéré encore plus facile que le niveau préconisé pour le style clair. Cela est sans doute dû au fait que le score de lisibilité est adapté au français, dont les mots sont généralement 1,5 fois plus longs qu'en français. Cependant, on peut souligner la difficulté et l'ambiguïté de certains éléments lexicaux notamment, qui font là aussi douter de la possibilité pour un.e lecteur.ice de 11 ans (âge pour lequel ce texte est facilement compréhensible d'après l'échelle), par exemple le verbe *nuit*, morphologiquement ambigu avec le nom de la même forme, et rarement employé dans sa forme verbale dans un registre courant ou familier. Des précautions sont donc de mise dans l'interprétation des scores FRE, comme nous en discutons plus bas.

Qu'en est-il de la lisibilité dans le corpus LEX, d'après les scores de FRE ? Tout d'abord que pour les deux corpus, certains textes obtiennent des scores FRE négatifs : respectivement neuf textes du corpus EN-LEX et huit du corpus EN-PLAIN. Pour le corpus EN-LEX, tous ces scores négatifs sont inférieurs à -100. Or, si l'échelle du score FRE Ease n'a pas de limite basse, c'est-à-dire qu'il est possible d'obtenir un score négatif, il est surprenant d'obtenir des scores aussi bas : le score le plus bas est de -570. La présence de points aberrants, c'est-à-dire très éloignés de la distribution majoritaire, interroge le calcul fait par l'outil de mesure sur les textes : un score de -500, sur une échelle censée aller de 0 à 100, même si les scores négatifs sont recevables, est-il encore pertinent ? La plupart des textes sont cependant distribués autour de 34. Un texte obtenant un score compris entre 0 et 30 est, selon Flesch (1979), considéré comme facile pour un.e *college graduate*, c'est-à-dire une personne ayant terminé un cursus universitaire standard. Flesch (1979) donne l'exemple d'un extrait de la revue de droit *Harvard Law Review*, qui obtient un score de 32, et un extrait du *Internal Revenue Code* qui obtient un score de -6. Greene et al. (2012), dans leur étude de la compréhension de plusieurs genres discursifs juridiques en anglais avant et après simplification, trouvent que sept textes

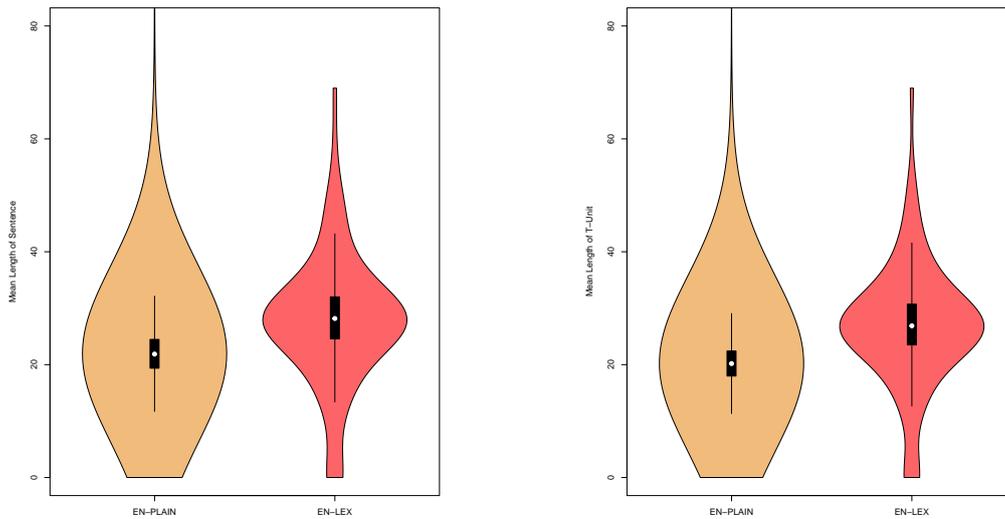
avant simplification obtiennent un score FRE compris entre 0,0 et 42,8 sur l'échelle, avec une moyenne de 7,4. Or, selon nos résultats, la moitié des textes qui composent le corpus EN-LEX obtiennent un score compris entre 0,0 et 34 (le score médian), sauf neuf textes qui obtiennent un score négatif. Ces résultats sont donc cohérents avec ceux obtenus dans d'autres études mesurant ce même score de lisibilité sur des textes juridiques. Onze textes sur 83 obtiennent en revanche un score au-dessus de 50, seuil qui correspond selon l'échelle à un lectorat d'élèves de cycle secondaire (lycée). Il y a donc dans ce corpus également encore une forme de variabilité entre les textes.

Pour le corpus spécialisé francophone FR-LEX, le score FRE médian est de 11. Comme nous l'avons vu plus haut, selon le lectorat modèle des mesures de Flesch, cela correspond à un lectorat de *college graduate*. Or, la classe FR-LEX est également caractérisée d'après le modèle par trois autres métriques de lisibilité : l'Automated Readability Index, qui a pour médiane 40, 21 pour le Linsear Write Formula et 32 pour la Flesch Kincaid Grade. Cela signifie qu'il faudrait entre 20 et 40 années de scolarité pour comprendre la moitié des textes du corpus FR-LEX ! De même plusieurs textes juridiques francophones obtiennent des scores négatifs, dont le minimum de -1449. Ces scores sont, nous l'avons vu, hors échelle, c'est-à-dire plus bas encore que les scores autour de 0 considérés comme très difficiles (*very difficult*) pour le lectorat visé, tandis que le sous-corpus FR-PLAIN est plus homogène. Là encore se pose la question de la pertinence d'un tel score et de problèmes de calcul de l'outil. Comment expliquer de tels scores de lisibilité ?

6.2.2 Longueur moyenne des phrases et des unités syntaxiques dans les corpus LEX et PLAIN

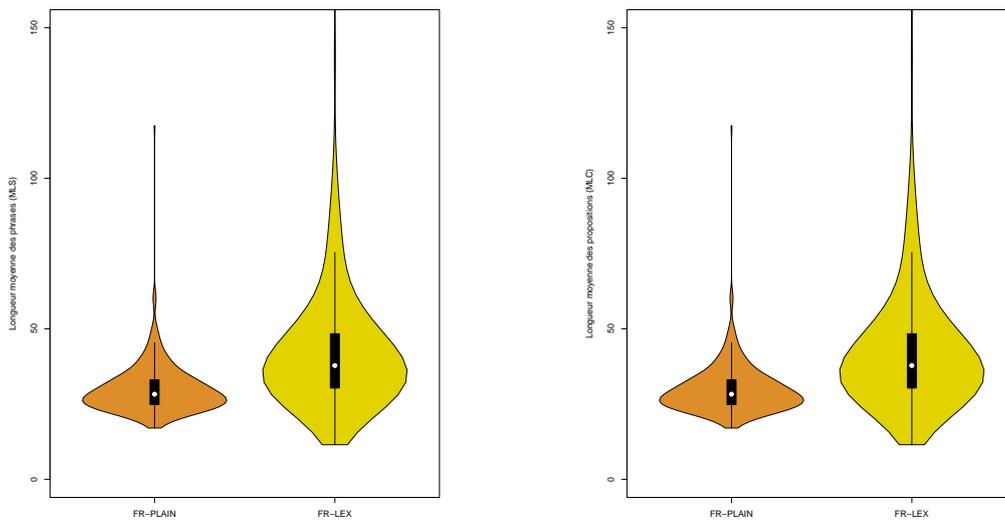
Les scores de lisibilité étant fondés sur des paramètres mesurant la longueur mots et des phrases, on peut se demander quelle est le paramètre qui joue le plus sur le score, et qui pourrait expliquer certains scores très bas. Selon le modèle de classification, la longueur moyenne des phrases est sélectionnée comme une métrique ayant un fort pouvoir de prédiction pour les deux langues ; c'est également le cas de la longueur moyenne des propositions (MLC) pour les textes anglophones et des T-Units (MLT) pour les textes francophones. Nous représentons ces métriques dans les Figures 6.6 et 6.7.

Pour les deux langues, on voit que la longueur moyenne des phrases est très proche de celle des T-Units ou des propositions. En regardant plus en détail, il



(a) Longueur moyenne des phrases dans EN-PLAIN et EN-LEX (b) Longueur moyenne des T-Units dans les corpus EN-PLAIN et EN-LEX

FIGURE 6.6 – MLS et MLT dans les corpus anglophones



(a) Longueur moyenne des phrases dans FR-PLAIN et FR-LEX (b) Longueur moyenne des propositions dans les corpus FR-PLAIN et FR-LEX

FIGURE 6.7 – MLS et MLC dans les corpus francophones

apparaît que la longueur moyenne des phrases du corpus EN-PLAIN correspond globalement aux recommandations pour un style clair, qui préconisent de limiter le nombre moyen de mots par phrase, avec en moyenne 23 mots par phrase. Elle est légèrement supérieure pour le français, avec 29 mots, mais la longueur des phrases est en moyenne plus longue pour cette langue (François et Fairon 2013). Globalement,

à nouveau les résultats pour les corpus LEX montrent davantage de points dans les valeurs hautes, et un score médian plus élevé (32 mots en anglais), tandis que la longueur moyenne des phrases reste autour de 23 et 28 mots respectivement dans EN-PLAIN et FR-PLAIN. On trouve dans ces corpus des textes ayant une longueur moyenne de phrase très haute, respectivement de 68 et 117. Dans le corpus spécialisé, la MLS monte à 69 pour l'anglais et 423 (!) pour le français.

Un retour aux contextes permet de distinguer plusieurs types de phrases dans le corpus LEX comme dans le corpus PLAIN. Dans les deux corpus, on trouve tout d'abord des phrases très longues, qui contiennent jusqu'à deux fois plus de mots que la limite supérieure des recommandations. On trouve de tels exemples dans les corpus EN-LEX et FR-LEX.

- (50) In this case, and // as I have explained, the first stage of the modified RJR-MacDonald test required the Crown // to satisfy the chambers judge // that there was a strong likelihood on the law and the [evidence presented] that it would be successful [in proving CBC's guilt of criminal contempt of court]. (52 mots, EN-Jugements)
- (51) An adoption order may be made on the application of one person // who has attained the age of 21 years // if the court is satisfied // that the person is the partner of a parent of the person // to be adopted. (40 mots, EN-Lois)
- (52) Le cadre d'analyse [énoncé dans Lessard], [qui vise [à mettre en balance] l'intérêt de l'État [à enquêter sur les crimes] et [à poursuivre] leurs auteurs et le droit des médias à la confidentialité des renseignements dans le processus de collecte et de diffusion des informations, // dresse une liste de neuf facteurs [dont les juges doivent tenir compte] lorsqu'ils décident // s'ils doivent délivrer un mandat de perquisition en lien avec les médias. (FR-CA-Jugements)

Les exemples (50) à (52), extraits de textes spécialisés, sont des phrases caractérisées par un haut degré de subordination et d'enchâssement, en particulier par la présence d'une ou plusieurs propositions complétives et relatives au sein d'une même phrase. Dans l'exemple tiré d'EN-Lois, la proposition relative post-modifie le nom-tête du GN, permettant ainsi de préciser l'âge de la personne concernée par la procédure juridique, tandis que la proposition conditionnelle a pour fonction d'indiquer le rôle de l'institution judiciaire. Pour certains des exemples, le haut degré d'imbrication syntaxique s'explique par la nécessité de préciser les conditions d'application de la loi, d'où la présence de propositions conditionnelles en *if*, *whether* ou *si*. On note également l'emploi de groupes nominaux postmodifiés par des groupes prépositionnels en chaîne (en particulier dans l'exemple 52 *le droit des médias à la confidentialité des renseignements dans le processus de collecte et de diffusion des informations*), qui allongent considérablement les propositions.

On trouve également des phrases exceptionnellement longues dans les corpus EN-PLAIN et FR-PLAIN.

- (53) If, [after exhausting the provider’s or local authority’s complaints procedure,] you are not satisfied with the outcome of your complaint, // you can complain to the Local Government Ombudsman // who can investigate complaints about local authorities in a fair and independent way. (41 mots, EN-Brochures)
- (54) It’s therefore worth checking // whether you may be entitled to other benefits // where you don’t need to show [that you have the right to reside], // although you do have to show [that you are habitually resident to claim some of these benefits]. (42 mots, EN-CAB)
- (55) Pour récupérer une dette impayée, // vous pouvez vous adresser à un huissier // pour lancer une procédure de recouvrement de créances, // ou demander gratuitement à un juge // de prononcer une injonction [de payer] à l’encontre de votre débiteur. (37 mots, FR-DILA)

Dans les exemples tirés des corpus PLAIN, les propositions subordonnées permettent d’élaborer pour indiquer au justiciable si sa situation personnelle est une des situations requérant le recours à la procédure judiciaire ou administrative en question. C’est pourquoi on trouve également des propositions subordonnées conditionnelles ou coordonnées expliquant les différentes démarches possibles (FR-DILA *vous pouvez vous adresser à un huissier...ou demander à un juge*). Les extraits des sous-corpus EN-CAB et EN-Brochures (53 et 54) contiennent quatre à cinq propositions dans une seule phrase, dont des propositions relatives et des propositions conditionnelles en *if* ou *whether*, ce qui illustre le grand nombre de propositions et le haut degré de subordination par phrase, d’où l’importance des métriques Mean Length of T-units et Clauses. Ces exemples tirés des corpus montrent qu’il y a dans chacun d’eux des phrases occasionnellement très longues, qui violent les recommandations pour un style clair, puisqu’elles demanderaient - toujours selon cette conception de la complexité - un plus grand effort cognitif aux lecteurs et lectrices. Certaines de ces phrases pourraient probablement être coupées en deux, mais puisque c’est la longueur moyenne des phrases qui doit être au-dessous de 25, il est logique que certaines phrases contiennent plus de mots que cette valeur.

C’est surtout un second type de phrases apparaissant dans les corpus qui semble pouvoir expliquer que les métriques liées à la longueur des phrases et des propositions ou T-Units ont été repérées par le modèle. Les scores FRE relativement bas de certains textes, et la MLS haute des textes juridiques et de certains textes PLAIN en particulier peuvent s’expliquer davantage par une autre caractéristique de ces textes, qui n’est pas prise en compte par l’outil de calcul des scores de lisibilité : la présence de listes à puces. Les textes législatifs, en particulier, qui sont divisés en sections et sous-sections, contiennent des énumérations énonçant les dispositions sur les cas et situations auxquelles s’applique la loi et les indications sur l’exécution de la loi. Nous présentons ci-dessous deux extraits des textes ayant obtenu des scores

FRE les plus bas.

- (56) The matters to which the Secretary of State is to have regard in determining for the purpose of subsection 1 whether the length of any delay is acceptable include the patient's medical history, the extent of any pain, disability, discomfort or other suffering that is attributable to the medical condition to which the service is to relate, whether any such pain, disability, discomfort or suffering makes it impossible or extremely difficult for the patient to carry out ordinary daily tasks, and the extent to which the provision of the service would be likely to alleviate, or enable the alleviation of, the pain, disability, discomfort or suffering. (EN-Lois)
- (57) Les peines sont portées à cinq ans d'emprisonnement et à 75 000 euros d'amende lorsque :
 Le conducteur a commis une violation manifestement délibérée d'une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement autre que celles mentionnées ci-après ;
 Le conducteur se trouvait en état d'ivresse manifeste ou était sous l'empire d'un état alcoolique caractérisé par une concentration d'alcool dans le sang ou dans l'air expiré égale ou supérieure aux taux fixés par les dispositions législatives ou réglementaires du code de la route, ou a refusé de se soumettre aux vérifications prévues par ce code et destinées à établir l'existence d'un état alcoolique ;
 Il résulte d'une analyse sanguine ou salivaire que le conducteur avait fait usage de substances ou de plantes classées comme stupéfiants, ou a refusé de se soumettre aux vérifications prévues par le code de la route destinées à établir s'il conduisait en ayant fait usage de stupéfiants ;
 Le conducteur n'était pas titulaire du permis de conduire exigé par la loi ou le règlement ou son permis avait été annulé, invalidé, suspendu ou retenu ;
 Le conducteur a commis un dépassement de la vitesse maximale autorisée égal ou supérieur à 50 km/h ;
 Le conducteur, sachant qu'il vient de causer ou d'occasionner un accident, ne s'est pas arrêté et a tenté ainsi d'échapper à la responsabilité pénale ou civile qu'il peut encourir. (FR-Lois)

Ces énumérations, sous forme de liste, forment les dispositions qui permettent aux législateur.ices d'indiquer avec le plus de détails possible le champs d'application ou d'exécution de la loi, comme on le voit dans ces deux exemples, où le premier établit la liste des éléments à prendre en compte pour des patient.es, et le second les circonstances aggravantes des peines liées à l'atteinte involontaire à l'intégrité physique lors d'une infraction en voiture. Il est notable que ces listes sont numérotées dans les textes législatifs originaux, comme le montre la Figure 6.8. Ces numéros ont été supprimés lors de la préparation des corpus pour l'analyse. Or, pour l'outil Textstat, une phrase est définie comme commençant par une majuscule et se terminant par un point. C'est également ainsi que Flesch indique de compter les phrases : *Disregard paragraph breaks, colons, semicolons, dashes or initial capitals within a sentence* (Flesch 1979). Les exemples ci-dessus sont donc considérés,

FIGURE 6.8 – Énumération dans un texte de FR-LEX

Les peines sont portées à cinq ans d'emprisonnement et à 75 000 euros d'amende lorsque :

- 1° Le conducteur a commis une violation manifestement délibérée d'une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement autre que celles mentionnées ci-après ;
- 2° Le conducteur se trouvait en état d'ivresse manifeste ou était sous l'empire d'un état alcoolique caractérisé par une concentration d'alcool dans le sang ou dans l'air expiré égale ou supérieure aux taux fixés par les dispositions législatives ou réglementaires du code de la route, ou a refusé de se soumettre aux vérifications prévues par ce code et destinées à établir l'existence d'un état alcoolique ;
- 3° Il résulte d'une analyse sanguine ou salivaire que le conducteur avait fait usage de substances ou de plantes classées comme stupéfiants, ou a refusé de se soumettre aux vérifications prévues par le code de la route destinées à établir s'il conduisait en ayant fait usage de stupéfiants ;
- 4° Le conducteur n'était pas titulaire du permis de conduire exigé par la loi ou le règlement ou son permis avait été annulé, invalidé, suspendu ou retenu ;
- 5° Le conducteur a commis un dépassement de la vitesse maximale autorisée égal ou supérieur à 50 km/h ;
- 6° Le conducteur, sachant qu'il vient de causer ou d'occasionner un accident, ne s'est pas arrêté et a tenté ainsi d'échapper à la responsabilité pénale ou civile qu'il peut encourir.

d'après ce calcul, comme étant une seule phrase de et une de 232 mots. Les textes législatifs comprenant souvent plusieurs listes de ce type, on comprend que la longueur moyenne des phrases soient en moyenne très élevées pour les corpus FR-LEX et EN-LEX, et que les scores de lisibilité qui se fondent dessus donnent des résultats suggérant une mauvaise lisibilité d'après les échelles utilisées. On comprend mieux par exemple pourquoi la médiane du score FRE pour FR-LEX est de 11.

Or, les textes procéduraux, c'est-à-dire ceux qui visent à guider les justiciables dans les démarches administratives ou juridiques, contiennent également des énumérations de ce type. Nous présentons des extraits de textes qui ont obtenu des scores au-dessus de la moyenne pour les métriques MLS et MLT, dans les corpus FR-PLAIN et EN-PLAIN.

(58) Une fois condamné, Vous pouvez

- demander un aménagement de peine sous écrou (semi-liberté, placement à l'extérieur, placement sous surveillance électronique) si vous avez été condamné à une ou plusieurs peines d'une durée totale inférieure ou égale à deux ans (un an en cas de récidive légale), ou si votre reliquat de peine d'emprisonnement restant à exécuter est inférieur ou égal à deux ans (un an en cas de récidive légale) ;
- demander une libération conditionnelle dès que vous avez effectué la moitié de votre peine ;
- rencontrer un personnel du service pénitentiaire d'insertion et de probation (SPIP) pour préparer un projet d'aménagement de peine. Le SPIP peut également vous aider pour une demande de suspension ou de fractionnement de votre peine ;
- sous certaines conditions, faire la demande d'une permission de sortir. (FR-Brochures)

(59) The costs of taking out income protection insurance are affected by :

- your age - the older you are when you take out the policy, the more you are likely to pay, as your risk of getting ill increases

- your sex - men make slightly more claims than women, so may pay more
- your health - if you're in good health, you will pay less to insure yourself
- your job - if you do a risky job, you will pay more for cover
- hobbies and lifestyle - if you take part in dangerous hobbies or you smoke or drink heavily, you will pay more for cover
- the waiting period - the longer you can wait before you make a claim, the cheaper your premiums will be
- whether you might be prepared to do other kinds of work than your own if you get ill - it usually costs less to take out income protection
- insurance if you say you will only make a claim if you are unable to do any work at all, rather than just your own job.

(EN-CAB)

Le premier extrait est issu d'une brochure destinée aux personnes en détention en France. La phrase, délimitée par la majuscule et le point, compte 162 mots. Cependant, l'énumération dans le texte lui-même est mise en page, conformément aux recommandations pour un style clair, sous forme d'une liste à puces, comme on le voit dans la Figure 6.9.

FIGURE 6.9 – Mise en page d'un texte du corpus FR-PLAIN

Une fois condamné,

■ Vous pouvez

- demander un aménagement de peine sous écrou (semi-liberté, placement à l'extérieur, placement sous surveillance électronique) si vous avez été condamné à une ou plusieurs peines d'une durée totale inférieure ou égale à deux ans (un an en cas de récidive légale), ou si votre reliquat de peine d'emprisonnement restant à exécuter est inférieur ou égal à deux ans (un an en cas de récidive légale) ;
- demander une libération conditionnelle dès que vous avez effectué la moitié de votre peine ;
- rencontrer un personnel du service pénitentiaire d'insertion et de probation (SPIP) pour préparer un projet d'aménagement de peine. Le SPIP peut également vous aider pour une demande de suspension ou de fractionnement de votre peine ;
- sous certaines conditions, faire la demande d'une permission de sortir.

Le second extrait est également considéré comme une très longue phrase, or il est clair qu'il utilise principalement des monosyllabes. Ces textes paraissent rester proches des textes législatifs en termes de structure, puisqu'ils semblent imiter cette stratégie discursive qui consiste à utiliser des listes pour présenter avec le plus de précision possible les différentes issues pour les justiciables, ou les différents facteurs qui entrent en jeu dans des décisions administratives, comme on le voit ci-dessus. Mais ces exemples montrent également que les scores de complexité syntaxique et de lisibilité reposent sur des métriques quantitatives qui ne prennent pas en compte d'autres caractéristiques pragmatiques ou multimodales des textes, puisqu'ils les

voient uniquement comme des chaînes de caractères. Les textes publiés par les institutions, en revanche, utilisent la mise en page pour fragmenter l'information en morceaux visuellement différenciés les uns des autres et faciliter la lecture, comme préconisé dans les recommandations en style clair.

Enfin, la présence de listes dans les textes procéduraux peut expliquer en partie la différence entre les scores FRE entre les sous-corpus de textes procéduraux et de Résumés. Ces derniers ne contiennent pas de listes de ce genre, ce qui peut expliquer pourquoi leurs scores FRE restent relativement stables et ne descendent pas en-dessous d'un certain niveau, alors que certains textes des sous-corpus EN-CAB, FR-DILA, mais surtout EN-Brochures et FR-Brochures, obtiennent des scores négatifs, comme nous l'avons vu dans les résultats quantitatifs (6.1.4.). Cela serait dû à la présence de listes telles que présentées ci-dessus, qui auraient une incidence sur la longueur moyenne des phrases, et donc sur le score FRE.

6.3 Analyse qualitative des métriques syntaxiques et lexicales ayant un fort pouvoir de prédiction

Les résultats suggèrent que certaines métriques de complexité syntaxique calculées par le *Syntactic Complexity Analyzer* ainsi que la série de micro-systèmes ont également un fort pouvoir de prédiction pour les textes anglophones, tandis que plusieurs métriques syntaxiques et lexicales ont un fort pouvoir de prédiction pour les corpus FR-LEX et FR-PLAIN.

6.3.1 Pour l'anglais : des métriques qui suggèrent la passage d'un style nominal à un style verbal

Le modèle de classification des textes assigne, pour la classe EN-PLAIN, une grande importance aux métriques portant sur le nombre de propositions (*Clauses*), de groupes verbaux (*Verb Phrases*) et, nous l'avons vu, la taille des T-units et des phrases, c'est-à-dire à des métriques liées au verbal et à la proposition. Les métriques qui permettent de distinguer le corpus EN-LEX, d'après le modèle, sont à l'inverse résolument associées au domaine nominal : groupes nominaux complexes (CN1), nombre d'articles définis (*Article THE*) et GN complexes contenant une préposition OF ou FOR (N_PREP_N). Comment ces différences sont-elles réalisées

linguistiquement et en contexte ? Comparons d'abord ces deux passages portant sur le droit syndical au Royaume-Uni, l'un issu du discours législatif, l'autre du discours de diffusion du droit Citizens Advice Bureau (CAB).

- (60) Everyone has the right to **freedom of peaceful assembly** and to **freedom of association with others**, including the right // to form and to join **trade unions for the protection of his interests**. No restrictions shall be placed on **the exercise of these rights** other than // such as are prescribed by law and // are necessary in a democratic society **in the interests of national security or public safety**. (EN-Lois)
- (61) If you 're a member of a trade union, you **should contact** them // to see if they **can help** with your problem at work, // before **trying** // **to find help** elsewhere. (EN-CAB)

Le premier exemple est issu du texte législatif *Human Rights Act 1998* et offre une illustration claire des caractéristiques capturées par les métriques, et associées au discours législatif. L'extrait, composé de deux phrases, contient six groupes nominaux complexes formés de structures de la forme <GN1 + OF/FOR + GN2> et même d'un enchaînement de deux GN en FOR puis OF. Le goût du style législatif pour les nominalisations est ici évident notamment avec l'utilisation de la nominalisation *for the protection of his [sic] interests* plutôt que de la proposition infinitive *to protect his interests*. Cet exemple montre bien en quoi ce sont les GN complexes pré et post-modifiés qui sont favorisés et donc capturés par les métriques. L'emploi de ces GN complexes déclenche grammaticalement l'utilisation de l'article défini, il est donc cohérent que la métrique comptabilisant THE ait été très prédictive. La loi énonce et définit, à l'aide de ces chaînes nominales, la liberté syndicale en tant que principe juridique abstrait. Dans le passage tiré de EN-CAB, à l'inverse, les rédacteur.ices viennent expliciter très concrètement une suite d'actions possibles, afin de guider les justiciables pour engager des démarches administratives et/ou des poursuites judiciaires. Pour cela, les textes contiennent de nombreux verbes d'actions (*should contact* ici) ainsi que des séquences verbales complexes, appelées chaînes verbales (*chain verbs*, notre traduction) par Halliday et Matthiessen (2013), comme ici *trying + to find help*. Ces résultats et analyses en contexte suggèrent donc que les textes du corpus EN-PLAIN présentent en réalité une forme de complexité verbale capturée par les métriques de complexité syntaxique, par opposition aux textes EN-LEX, qui sont caractérisées par des métriques liées à la complexité nominale. Cette hypothèse semble confirmée par les exemples suivants, cette fois extraits du discours juridictionnel et de sa reformulation pour le grand public.

- (62) The integrity of the administration of justice depends upon the unique role of the solicitor // who provides legal advice to clients within this complex system. (EN-CA-Jugements)
- (63) Accurate information about an offender's psychological needs and the risk // he or she poses // is crucial to achieving the system's purpose of [contributing to the maintenance of a just, peaceful and safe society] by carrying out sentences through safe and humane custody of inmates and // assisting in their rehabilitation and reintegration into the community. (EN-CA-Jugements)
- (64) Everyone **has a right** to the strongest protections // **to make sure** // they **aren't held** in custody against the law, // the Supreme Court **has ruled**. (EN-CA-Résumés)
- (65) Privilege protects certain kinds of information from being revealed. "Solicitor-client privilege" is a well-known type, // and **protects** communications between a client and a lawyer. // But there are other kinds.// This case **dealt with** "informer privilege,"// which **protects** the identities of confidential informers who help police. (EN-CA-Résumés)

Les exemples 62 et 63, extrait de deux jugements de la CS du Canada, illustrent à nouveau la préférence du discours juridique pour le style nominal, avec des GN post-modifiés par des groupes prépositionnels qui forment des cascades de noms, décrites par les défenseur.euses du style clair - on notera en particulier l'enchaînement d'un GN1 OF GN2 et de la nominalisation dans l'extrait 63, avec *the system's purpose of contributing to the maintenance of a just, peaceful and safe society*. Ces groupes nominaux en chaîne, comportant de nombreuses modifications adjectivales et prépositionnelles, permettent de faire tenir de manière compact, sous forme nominale. Ces GN participent de l'énoncé de principes sur lesquels se fondent les systèmes judiciaire et pénitentiaire (dans nos exemples), des connaissances partagées par les membres de la communauté spécialisée mais évoquée par les juges de la CS pour rappeler les fondement juridiques sur lesquels les cas reposent. Dans les exemples 64 et 65, à l'inverse, les extraits des résumés de jugements de la CS sont caractérisés non pas par des structures nominales complexes, mais par des structures verbales complexes. Dans l'exemple 64, on trouve d'abord les prédicats complexes lexicalisés comme *has a right* ou *make sure* qui entraîne l'utilisation d'une proposition, qui contient par ailleurs un passif, ce qui offre un bel exemple d'*intricacy*. Cette phrase contient par ailleurs une quatrième proposition "*the Supreme Court has ruled*", décrite comme *projecting clause*, c'est-à-dire qui introduit du discours indirect. Or ici, la proposition en *that* est antéposée, contrairement à une structure plus typique [verbe de parole + proposition nominale en *that*]. La métrique Clause, qui capture le nombre de propositions, est illustrée ici et rappelle que les résumés de jugements sont caractérisés par une forme d'hétérogénéité énonciative explicite (Preite 2018), puisqu'ils rapportent la parole des expert.es du droit, ici les juges de la Cour. Le second extrait de résumés, 65, exemplifie comment les rédacteur.ices explicitent les termes

spécialisés, à l'aide de propositions coordonnées (*and protects communications between a client and a lawyer* ou subordonnées relatives ; cette complexité syntaxique sert donc la ré-élaboration des connaissances juridiques pour les expliciter et les rendre accessibles aux justiciables.

Dans le corpus anglophone, le passage au PL et la reformulation du discours juridique passeraient donc paradoxalement par une plus grande complexité propositionnelle (*clausal complexity*, qui se traduit quantitativement par les métriques liées aux GV et proposition, bien que les phrases et propositions soient en moyenne moins longues, elles sont caractérisées quantitativement et qualitativement par une plus grande complexité verbale. Ces résultats, qui suggèrent une plus grande complexité syntaxique au niveau de la proposition (*intricacy*, sont cohérents avec ceux d'autres études du PL :

It is thus notable how much more intricate the so-called 'simplified' discourse is in terms of syntactic and morphological structure. (Gledhill et al. 2019)

Ces résultats, qui correspondent aux préconisations pour un style clair, sont aussi en miroir des caractéristiques du discours juridique, quant à lui souvent associé à un style nominal, ce qui est confirmé ici par le modèle d'apprentissage supervisé. Les *features* syntaxiques ayant le plus grand pouvoir de prédiction le modèle sont donc cohérent à la fois avec la littérature sur le discours du droit et avec les guides rédactionnels pour un style clair, ce qui reflète une même conception de la complexité linguistique.

6.3.2 Pour le français : des différences syntaxiques mais une même complexité lexicale ?

Deux métriques syntaxiques différentes caractérisent les corpus FR-LEX et FR-PLAIN d'après le modèle. Pour le corpus FR-LEX, la métrique est CP (*Coordinate Phrases*), soit la métrique qui mesure le nombre de coordinations. La métrique inclut en effet non seulement les GN coordonnées, mais aussi les groupes verbaux, adjectivaux et adverbiaux coordonnées, dont nous voyons des exemples ci-dessous, issus d'un jugement de la CS du Canada.

- (66) Les effets dissuasifs ne se prêtent pas à **une preuve scientifique ou empirique**. Ils sont plutôt de nature **intangibles et difficiles** — voire impossibles — à mesurer avec exactitude. Cela ne signifie pas qu'ils sont **insignifiants et sans importance**. Au contraire, dans la mesure où des effets dissuasifs se produisaient **bel et bien**, leurs conséquences pourraient être considérables : une presse **faible et craintive**, épaulée par un bassin de sources de

moins en moins important, se traduit par une société **moins informée, moins ouverte et moins dynamique** dans laquelle **la discussion, le débat et le flux de l'information** sont entravés. (FR-CA-Jugements)

La longueur moyenne des phrases étant une métrique assignée à ce corpus, il n'est pas étonnant que la coordination en soit également une, puisque nous l'avons vu les phrases longues contiennent des énumérations souvent coordonnées. L'exemple ci-dessus rappelle également l'affection du milieu juridique pour les expressions binomiales et trinomiales, souvent citées comme caractéristiques du discours du droit (Goźdz-Roszkowski et Pontrandolfo 2017), ici illustrées en gras. Ces groupes adjectivaux et nominaux coordonnés rappellent la volonté d'être le plus précis possible pour les législateur.ices et les magistrat.es dans la définition du domaine d'application de la loi et dans les motifs de jugements.

Concernant le corpus FR-PLAIN, le modèle suggère qu'il est caractérisé par la métrique VP correspondant au nombre de groupes verbaux. Dans l'exemple ci-dessous, on compare un extrait du code pénal français et un extrait du corpus FR-DILA expliquant le déroulement d'une perquisition. Les groupes verbaux (GV) sont présentés en gras.

- (67) Si la perquisition **a lieu** dans un domicile autre que celui de la personne mise en examen, la personne chez laquelle elle doit **s'effectuer est invitée** à y assister. Si cette personne **est** absente ou **refuse** d'y assister, la perquisition a lieu en présence de deux de ses parents ou alliés présents sur les lieux, ou à défaut, en présence de deux témoins.
- (68) - Si l'occupant **est** déjà mis en examen : il **doit être** présent ou **désigner** un représentant. Sinon, l'OPJ¹ **désigne** 2 témoins.
 - Si l'occupant **n'est pas mis** en examen, il **peut être** présent. Sinon, l'OPJ **désigne** 2 membres majeurs de la famille présents sur place. S'il n'y **a** personne, l'OPJ **désigne** 2 témoins majeurs (hors policiers ou gendarmes). (FR-DILA)

On voit que la préposition complexe "en présence de + GN", qui apparaît deux fois dans l'extrait du discours législatif pour spécifier qui peut être témoin d'une perquisition, est dans l'extrait destiné au grand public reformulé à l'aide de la réintroduction d'un référent humain et l'explicitation du rôle de l'Officier de Police Judiciaire, sujet répété du verbe *désigner* dans les différents cas de figure. D'une phrase contenant des groupes nominaux complexes indiquant des cas de figures abstraits et impersonnels, on passe à plusieurs phrases contenant des verbes conjugués, décrivant très concrètement les différentes possibilités et incluant une *repersonnalisation* de la situation décrite. La reformulation du discours juridique passe donc aussi en français, il semblerait, par le passage d'un style nominal à un style verbal

1. Officier de Police Judiciaire

à travers l'emploi de davantage de groupes verbaux, capturé par cette métrique de complexité verbale. Ces résultats sont là encore en accord avec les recommandations pour un style clair, les verbes étant considérés, nous l'avons vu, comme plus concrets que les noms (Collette et al. 2002).

Les deux corpus francophones sont par ailleurs caractérisés par deux métriques de complexité lexicale proches. Celle qui caractérise le corpus FR-PLAIN d'après le modèle est la métrique *Sophisticated Word Types*. Pour le corpus FR-LEX, la métrique lexicale est intitulée *Sophisticated Lexical Types*. Ces deux métriques parentes se fondent sur une liste de mots en français incluses dans l'outil *Lexical Complexity Analyzer* adapté au français. Cette liste contient des mots (non lemmatisés, comprenant donc des verbes conjugués ou adjectifs avec inflexion pour le féminin ou le pluriel) considérés comme rares (*sophisticated*), c'est-à-dire ne se trouvant pas dans les 2000 mots les plus fréquents dans les corpus généraux d'après cette liste. Il est à noter que ces métriques se concentrent sur les *word types* et non les *word tokens*, c'est-à-dire que l'outil calcule un score fondé sur le nombre de mots "rares" dans les corpus en les comptant une seule fois, il ne prend donc pas en compte la répétition. Or, il apparaît d'après le modèle que ces métriques de complexité lexicale ont un fort pouvoir de prédiction à la fois pour le corpus FR-LEX et FR-PLAIN. Ci-dessous sont présentés plusieurs extraits de ces corpus, pour tenter d'illustrer ce résultat. Les mots lexicaux considérés comme rares d'après le LCA sont en gras.

- (69) L'**allégation** de discrimination découle non pas de la différence entre ces **échéanciers**, ce qui ne constitue pas l'élément central de la plainte, mais de l'étendue du délai découlant de l'échéancier applicable aux demanderesse. Soutenir que l'existence d'échéanciers différents signifie que l'un de ces groupes ne fait pas partie de la catégorie des "bénéficiaires visées" constitue, soit dit en tout respect, une interprétation **erronée** du régime législatif et de la notion de "discrimination à rebours". Nous devons maintenant examiner l'article premier de la Charte. Il **incombe** au Québec de justifier la violation *prima facie* du par.

Il est aisé, en lisant cet extrait, de comprendre pourquoi la métrique a été repérée par le modèle, puisque ce texte contient de nombreux termes comme *allégations* ou le très administratif *échéancier*, mais également des éléments lexicaux appartenant à un registre soutenu (et donc perçu comme tels par la métrique), comme ici la collocation *interprétation erronée*. Ces résultats bien sûr sont parfaitement concordants avec la description du discours juridique dans la littérature évoquée au Chapitre 1. Il peut paraître plus surprenant que le sous-corpus FR-PLAIN soit également caractérisé par une métrique de complexité syntaxique. Regardons quelques extraits de corpus :

- (70) La question principale en **litige** était de savoir si la **fiducie**, représentée par MM. Brunette et Maynard, pouvait poursuivre les comptables et les avocats. Avant le procès, ces derniers ont demandé au tribunal de rejeter la **poursuite**, soutenant que MM. Brunette et Maynard n'avaient **manifestement** aucun "intérêt" pour **intenter** celle-ci. (FR-CA-Résumés)
- (71) Cette affaire portait sur la responsabilité d'un avocat dans une situation où il va **au-delà** de la simple recommandation d'un autre professionnel à ses clients. (FR-CA-Résumés)

Ces exemples, tirés de FR-CA-Résumés, montrent que la reformulation du discours juridique emploie non seulement des termes, ici *fiducie* et *litige* en particulier, mais également des collocations spécialisées comme (*intenter une poursuite*, (70)), ainsi que des prépositions considérées comme rares, comme *au-delà* dans l'exemple (71) (d'après la liste du LCA). C'est à cause de ces prépositions notamment, considérées comme des mots grammaticaux, que la métrique est liée aux *word types* et non aux *lexical types*. Tout se passe comme si le discours dans le corpus FR-PLAIN exprimait non seulement la voix des rédacteur.ices s'adressant aux justiciables, mais devait également adopter la voix et le ton de l'institution juridique et administrative pour pouvoir expliquer précisément les procédures et concepts. La question de la possibilité de la simplification de la terminologie juridique, et de la fréquence des termes et mots lexicaux considérés complexes, est une question que nous abordons en détail au Chapitre 7. Il est cependant notable que les métriques pointent une forme de complexité lexicale à la fois pour FR-LEX et pour FR-PLAIN, bien que ce dernier corpus soit également caractérisé par une forme de complexité verbale (propositionnelle).

6.4 Discussion : apports et limites de l'approche quantitative

Après avoir décrit les résultats quantitatifs et illustré qualitativement ces résultats, nous proposons un retour critique sur notre approche quantitative en revenant plus spécifiquement sur certains problèmes méthodologiques brièvement évoqués tout au long de ce chapitre.

6.4.1 Problèmes méthodologiques dûs aux données

En premier lieu, l'une des limites méthodologiques qui mérite d'être soulignée dans cette étude provient tout d'abord à la structuration des données et de la constitution du corpus.

Tout d'abord, le jeu de données anglophone contient, nous l'avons déjà évoqué, un déséquilibre importante en termes de nombre de textes. Si le corpus EN-PLAIN contient plusieurs centaines de textes, le corpus EN-LEX n'en contient quant à lui que 82. Les textes législatifs et jugements rassemblés sont cependant des textes de plusieurs milliers de mots, c'est pourquoi le corpus contient près de deux millions de mots, soit plus de deux fois plus que dans le corpus EN-PLAIN. Cela peut introduire un biais dans le modèle d'apprentissage en faveur de la classe qui contient davantage de points de données (la classe EN-PLAIN, dans notre cas), ce qui explique sans doute que l'algorithme arrive aussi bien à prédire les classes pour l'anglais. Puisqu'il y a un nombre de points bien supérieur pour la catégorie EN-PLAIN, même en assignant aléatoirement les étiquettes aux textes, l'algorithme aurait de grandes chances de trouver la bonne étiquette des textes, pour ceux du corpus EN-PLAIN en particulier. La régression logistique permet cependant de corriger partiellement ce biais.

Dans le jeu de données francophone, à l'inverse, puisque les textes législatifs sont présents dans le corpus FR-LEX sous la forme d'articles de lois, et non rassemblés en un seul texte comme dans le corpus EN-LEX, il n'y a pas de déséquilibre et le nombre de textes utilisés pour tester l'algorithme est sensiblement le même pour FR-PLAIN et FR-LEX (autour de 400 textes). Or, comme indiqué dans la description de la performance du modèle, l'algorithme a des difficultés à trouver les textes de la classe FR-LEX en se fondant sur les métriques. Ce corpus étant composé de nombreux extraits de textes législatifs, dont certains sont très courts (moins de 20 mots), il est possible que certains de ces textes contiennent moins de marqueurs capturés par les métriques qui seraient considérés comme des marqueurs de type LEX par l'algorithme, et que leur *features* soient donc classés comme PLAIN. L'autre hypothèse serait que pour certains textes des corpus FR-LEX et FR-PLAIN, la différence entre les deux registres (spécialisé/non-spécialisé) est moins évidente pour diverses raisons, notamment lexicales. Cette complexité lexicale proche pourrait entraîner une confusion des deux classes pour l'algorithme.

Par ailleurs, une seconde limite que nous devons à garder à l'esprit concernant le calcul des métriques est liée à la variabilité de longueur des textes auquel est assigné

pour chaque texte un unique score. Par exemple, certains textes du sous-corpus EN-CA-Jugements sont très courts tandis que d'autres contiennent plusieurs dizaines de pages.

- (72) The judgment of the Court was delivered orally by Moldaver J. We are in essential agreement with the reasons of Justice O’Ferrall. In particular, we agree that the learned trial judge conflated the actus reus and mens rea of the offence and did not sufficiently explain the concept of marked departure in a way that the jury could understand and apply it. Accordingly, we would allow the appeals, quash the convictions and order a new trial. Judgment accordingly.

Le texte ci-dessus est reproduit en entier, car il s’agit d’un des jugements les plus courts de la CS du Canada. De même pour les corpus EN-PLAIN et FR-PLAIN : certains textes sont très longs, d’autres très courts, mais un score unique leur est assigné pour chaque métrique, score qui ne prend pas nécessairement en compte la variation de longueur des textes. Il est logique qu’un texte de plus de 10 000 mots contiennent de nombreux articles définis, par comparaison à un texte court comme celui présenté ci-dessus, ce qui explique la grande variabilité des scores pour les textes du corpus LEX. De même, les textes PLAIN et LEX sont bien sûr de longueur très différentes, ce qui a une incidence sur certaines mesures, notamment celles qui ne mesurent pas des ratios mais le nombre de formes elles-mêmes, comme c’est le cas pour plusieurs des métriques syntaxiques présentées plus haut (*Coordinate phrases*, *Verb phrases*). Bien que la régression logistique avec régularisation lasso neutralise en partie ce problème, il est important d’interpréter les résultats avec précautions.

6.4.2 Biais et erreurs dans les outils statistiques : peut-on faire confiance aux métriques de complexité ?

Les outils automatisés eux-mêmes qui ont été utilisés pour extraire les métriques ne sont par ailleurs pas exempts d’erreurs. Ainsi, plusieurs métriques issues de la version modifiée du L2SCA incluant les microsystemes, qui calculent des ratios de formes dans les textes, donnent des scores minimaux de 0 pour plusieurs textes. C’est le cas par exemple de plusieurs textes pour la métrique N_PREP_N qui, nous l’avons vu, est une métrique très prédictive pour la classe LEX. D’après les données, neuf textes du corpus EN-LEX obtiennent un score de 0.0 pour cette métrique, qui, nous le rappelons, mesure le nombre de GN complexes composé de deux groupes nominaux et de la préposition *for* ou *of*. Nous reproduisons ci-dessous un extrait d’un des textes du corpus EN-Lois, le *Immigration Act 2016* du Royaume-Uni.

- (73) In this section "labour abuse prevention officer" means **an officer of the Gangmasters and Labour Abuse Authority** who falls within subsection , and is authorised (whether generally or specifically) by **the Secretary of State for the purposes of this section**.

En gras apparaissent les GN complexes qui correspondent à la structure censée être repérée par cette métrique, notamment la fameuse séquence *for the purposes of this section*. Or, le score obtenu pour cette métrique est de 0, ce qui montre que l'outil a fait une erreur et n'a pas réussi à capturer les nombreuses structures correspondant. L'automatisation des mesures de complexité présente donc également des risques d'erreur, qui ne sont pas à écarter, puisqu'on peut se demander si certaines métriques n'ont pas obtenu les bons scores et auraient pu avoir un pouvoir de prédiction important. Il est par ailleurs difficile de déterminer d'où vient cette erreur de calcul : la métrique ayant été sélectionnée par le modèle, elle a nécessairement bien été calculée sur la plupart des textes de EN-PLAIN. La présence de points aberrants dans les données pose également question quant à la réussite des calculs faits par les différents outils dans la chaîne de traitement, même si dans ce cas la longueur des listes et énumérations dans les textes législatifs peut expliquer en partie ces scores hors échelle. En particulier, les scores FRE de -1449 et de 423 pour la longueur moyenne des phrases interroge l'adaptation de l'outil au français.

Bien sûr, la façon dont les métriques elles-même sont conceptualisées et implémentées a une incidence sur les mesures et peuvent entraîner des erreurs de classification. Le passage suivant, issu du sous-corpus FR-Lois, obtient ainsi un score FRE de 75, c'est-à-dire qu'il est censé être facile à comprendre pour un lectorat-modèle de 10 ou 11 ans.

- (74) L'intérêt est légal ou conventionnel. L'intérêt légal est fixé par la loi. L'intérêt conventionnel peut excéder celui de la loi, toutes les fois que la loi ne le prohibe pas. Le taux de l'intérêt conventionnel doit être fixé par écrit.

Ce court article de loi est effectivement composé de phrases de 10 mots ou moins, et de mots relativement courts (en particulier *loi*, *intérêt*) mais, on voit bien ici que les termes et collocations *intérêt conventionnel* ou le vocabulaire, notamment les mots *excéder* et *prohibe*, bien que composés de peu de syllabe, appartiennent à un registre soutenu. Le modèle d'ailleurs ne s'y trompe pas puisqu'il classe le texte dans la catégorie LEX, le texte contenant d'après les scores 19 mots lexicaux (*tokens*) considérés comme "sophistiqués" sur 19 mots lexicaux et sur 41 mots au total (mots

grammaticaux inclus). Mais les autres métriques ne suffisent pas toujours, et le poids du score FRE peut entraîner des erreurs de l'algorithme dans la classification. Comme nous l'avons vu, certains textes procéduraux du corpus PLAIN, c'est-à-dire les textes expliquant les procédures judiciaires et administratives, contiennent des listes pour préciser différentes conditions ou possibilités d'application de la loi. Les listes étant considérées comme une seule phrase par l'outil, nous l'avons vu, certains textes sont considérés comme comportant de très longues phrases, ce qui est perçu dans les métriques comme nuisant à la lisibilité. Ci-dessous est présenté un des textes EN-PLAIN catégorisés LEX par le modèle :

(75) Some people can't serve on a jury :

- anyone with an intellectual disability, as defined in section 7 of the Intellectual Disability (Compulsory Care & Rehabilitation) Act 2003
- a member of parliament the Governor-General
- a judge, community magistrate or visiting justice
- a member of the Parole Board a barrister or solicitor with a current practising certificate
- Justice of the Peace who hears cases in the District
- Court employees of the following organisations : Department of Corrections New Zealand Police Ministry of Justice.

Ce texte est constitué d'une longue énumération des personnes qui ne peuvent pas être convoquées comme juré.es pour un procès en Nouvelle-Zélande. Cette liste étant présentée sous la forme d'une liste à puces (avec une seconde liste insérée) sans ponctuation, est comptée par l'outil de mesure automatisée comme une phrase longue de plusieurs dizaines de mots. Le texte obtient donc un score FRE de -130, ou un Flesch-Kincaid score de 76 (c'est-à-dire qu'il faudrait 76 ans de scolarité pour le comprendre). Pour le modèle d'apprentissage (et à cause des autres métriques également), ce texte est donc classé dans la classe LEX, puisqu'il obtient l'un des scores FRE les plus bas du jeu de données. Selon la manière dont ces scores ont été conceptualisés, et également selon ce que les guides rédactionnels recommandent pour un style clair, cette phrase semble en effet complexe, puisque, pour être mémorisée, elle demanderait un effort cognitif important. Cependant, on voit d'une part que cette phrase est syntaxiquement très simple. D'autre part, si l'on revient à la définition du *plain language* tel qu'il est défini, c'est-à-dire que le ou la lectrice peut y trouver l'information qu'il recherche à la première lecture, il semble que cet extrait du corpus EN-Brochures permet bien à ses lecteurs ou lectrices de savoir si

oui ou non ils ou elles appartiennent aux catégories de personnes qui ne peuvent pas être appelées à siéger dans un jury.

A l'inverse, certains textes obtenant un score FRE correspondant à un niveau "facile" de lecture jettent le doute sur la cohérence des métriques entre elles. L'exemple suivant est extrait d'un texte du corpus FR-CA-Résumés destiné au grand public et qui a obtenu un score FRE de 63,6.

- (76) Selon la Cour suprême, l'infraction de trafic d'influence englobe la promesse d'influencer le gouvernement pour qu'il change son mode de fonctionnement. L'expression "un sujet d'affaires ayant trait au gouvernement" qui figure au Code criminel vise les activités qui vont au-delà des opérations gouvernementales en cours et s'étend à celles que le gouvernement pourrait faciliter. Dans une décision rendue à 8 voix contre une, la juge Andromache Karakatsanis a rejeté l'appel de Bruce Carson et confirmé sa déclaration de culpabilité pour trafic d'influence.

Ce texte a donc obtenu un score correspondant au style clair d'après les recommandations (un score de 60 ou au-dessus). La longueur moyenne des phrases est pourtant de 29 mots, et, d'après les autres métriques, ce texte contient 135 mots sophistiqués sur 140 mots lexicaux, et requiert entre 12 et 15 ans de scolarité d'après l'Automated Readability Index et la Linsear Write Formula pour être lisible. Les résultats du score FRE vis-à-vis des autres métriques ne sont donc pas toujours cohérents.

Ces résultats confirment certaines critiques qui ont été portées aux approches quantitatives et aux scores de lisibilité notamment. De par la façon dont ils ont été conceptualisés et implémentés dans les outils de mesure automatique, ils sont excessivement sensibles à certaines caractéristiques des textes. Ils reflètent également certains stéréotypes sur la complexité linguistique, notamment qu'une phrase longue est forcément complexe, et qu'un texte contenant des phrases courtes est aisé à comprendre. Enfin, ces métriques, puisqu'elles considèrent les textes comme des chaînes de mots, ignorent certaines caractéristiques pragmatiques centrales des textes, comme leur mise en page. Or, nous l'avons vu, une phrase peut contenir beaucoup de mots mais être en réalité structurellement très simple, et donner à son lecteur ou sa lectrice les informations qu'il ou elle recherche en un seul coup d'œil.

Pour autant, faudrait-il jeter le bébé avec l'eau du bain, et brûler les métriques de complexité, comme se le demandait Richaudeau pour les formules de lisibilité, dans son article au titre provocateur (Richaudeau 1976)? Ici nous arguons que nos résultats, malgré leurs limites, sont parfaitement cohérents avec la vision de la com-

plexité présentée dans la littérature sur le style clair Les métriques et l'algorithme de classification, bien qu'incomplets et comportant des biais nombreux offrent, donnent des indices précieux sur le passage du style juridique au style clair. Ils permettent de voir que les rédacteur.ices ayant pour mission de simplifier le discours du droit pour le rendre accessible respectent les recommandations en style clair, au moins partiellement, notamment dans la ré-élaboration des connaissances juridiques dans la proposition. Ces métriques permettent également de mettre au jour des différences possibles entre les genres discursifs (bien qu'une véritable comparaison entre les genres n'ait été que partiellement abordée ici), tout en interrogeant la conception de la complexité linguistique sur lesquelles elles reposent.

6.5 Résumé du Chapitre 6

Dans ce chapitre, nous avons vu que d'après le modèle d'apprentissage supervisé, les métriques de complexité lexicales, syntaxiques et de lisibilité peuvent être de bons indicateurs du passage de textes perçus comme complexes à des textes se réclamant du style clair. Ces résultats semblent confirmer la pertinence de certaines mesures classiques de lisibilité, comme le score FRE et la longueur des phrases, mais indiquent également que la reformulation du discours du droit passe par une plus grande complexité verbale et propositionnelle. Il est intéressant de constater pourtant que ce sont les métriques de complexité syntaxique, en particulier liées au verbal, qui sont les plus représentées dans les résultats anglophones, et partiellement pour les résultats francophones. Cependant, les résultats eux-mêmes doivent être envisagés avec de grandes précautions. Les métriques et la classification sont caractérisées par des erreurs à de multiples niveaux, dues à certains problèmes dans les outils automatiques mais aussi dans les données. Ces métriques sont donc de bons outils pour explorer le passage du corpus LEX au corpus PLAIN, à condition de systématiquement vérifier en contexte ce que ces métriques signalent au niveau linguistique. Cette analyse linguistique nécessite par ailleurs de prendre en compte d'autres éléments que ces seules métriques, en puisant notamment dans les outils plus classiques de la linguistique de corpus.

Chapitre 7

Approche lexicale et terminologique des textes juridiques et de leurs versions simplifiées

Dans ce chapitre, nous examinons le discours de diffusion du droit et du style clair sous l'angle des recommandations lexicales et terminologiques. Nous rapportons et analysons nos résultats concernant la fréquence des éléments lexicaux qualifiés de "pompeux" (Petelin 2010) et la densité terminologique dans les corpus EN-LEX et EN-PLAIN. Une attention particulière est ensuite portée aux stratégies métalinguistiques d'explicitation et de reformulation des termes dans les différents sous-corpus EN-PLAIN. Par exemple, nous nous demanderons si les stratégies de reformulation des mots non-approuvés, quand ils sont présents, mobilisent leur alternative claire et simple. Pour cette approche, nous nous concentrons sur les corpus anglophones EN-LEX et EN-PLAIN, dans lesquels nous recherchons la présence (ou non) de mots et expressions "non-approuvés" et de termes spécialisés. Tous les fichiers des sous-corpus ont été concaténés en un seul fichier texte pour faciliter le traitement. Nous obtenons ainsi la fréquence moyenne des mots à éviter selon les organismes promouvant le style clair, et de termes, dans chaque sous-corpus.

Pour le français, il convient de bien différencier les corpus issus de France et du Canada, en particulier pour l'étude de la densité terminologique. En effet, ces deux pays appartiennent à deux systèmes juridiques différents, l'un de droit civil, l'autre bijuridique (droit civil et *common law*). Ils ne relèvent donc pas des mêmes

systèmes et procédures judiciaires, ce qui est reflété dans l'emploi de terminologies très différentes. Pour les éléments lexicaux, les deux variétés sont étudiées, mais pour l'étude des termes, nous avons décidé de nous concentrer sur les corpus canadiens FR-CA-Jugements et FR-CA-Résumés.

7.1 Les textes anglophones se réclamant du style clair respectent-ils les normes lexicales ?

Dans cette section, nous présentons les résultats de l'étude du respect des prescriptions lexicales : si les recommandations sont respectées, on ne devrait trouver pas ou peu d'items lexicaux "non-approuvés" provenant de la liste des mots à éviter selon l'association Plain English Campaign.

7.1.1 Fréquence du vocabulaire "non-approuvé" : résultats quantitatifs

Nous présentons d'abord les résultats pour le corpus EN-LEX, en distinguant les deux sous-corpus dans le Tableau 7.1.

TABLEAU 7.1 – Proportion d'items lexicaux "non-approuvés" dans les sous-corpus spécialisés

Corpus	Lectorat visé	Proportion des items lexicaux "non-approuvés" (%)
EN-CA-Jugements	Expert	2,56
EN-Lois	Expert	4,35

La fréquence normalisée moyenne des items de la liste indique que les textes de lois contiennent plus d'items lexicaux à éviter que le corpus EN-CA-Jugements, ce dernier contenant 2,64% d'items lexicaux de la liste contre 4,35% du nombre total de *tokens* dans le corpus EN-Lois. Ces résultats sont intéressants en ce qu'ils suggèrent des différences à l'intérieur même du registre juridique, puisque les deux sous-corpus représentent deux sous-registres et deux genres différents. En moyenne, le corpus de textes législatifs contient donc davantage d'items lexicaux listés comme "à éviter" par l'association Plain English Campaign (PEC) : ils seraient donc, d'après ce résultat, plus complexes lexicalement que les textes issus du discours judiciaire. Il s'agit,

nous l'avons vu, d'une liste de mots complexes *a priori*, créée empiriquement par des éditeurs. Ce résultat peut cependant contribuer à expliquer la perception du discours législatif comme particulièrement étranger et comme le sous-registre du droit "le plus ésotérique" (Maley 1994) par les non-spécialistes. Ce caractère ésotérique pourrait donc en partie être imputé à l'utilisation de mots perçus comme complexes car appartenant à un registre soutenu, qualifié de pompeux (*fancy, pompous*) par les défenseurs.euses du style clair.

Le Tableau 7.2 présente quant à lui la fréquence moyenne globale de ces items lexicaux dans les sous-corpus du corpus EN-PLAIN. A première vue, les résultats

TABLEAU 7.2 – Proportion d'items lexicaux "non-approuvés" dans les sous-corpus destinés au lectorat non-expert

Corpus	Lectorat visé	Proportion de mots non-approuvés ¹ (%)
EN-CA-Résumés	Non-expert	0,89
EN-Brochures	Non-expert	2,10
EN-CAB	Non-expert	1,35

indiquent d'abord qu'il n'y a pas de conformité complète avec les règles lexicales prescrites pour rédiger en style clair : on trouve bien les mots non-approuvés, à des degrés divers dans chacun des sous corpus. Le sous-corpus EN-CA-Résumés est celui qui contient le moins d'éléments lexicaux jugés complexes, avec seulement 0,89% du nombre total de mots correspondant à des items lexicaux issus de la liste de mots éviter. C'est donc le corpus qui adhère le plus aux recommandations lexicales (toujours selon cette liste). A l'inverse, le corpus EN-Brochures est celui qui en contient le plus ; il est donc lexicalement plus complexe que les deux autres, si l'on se fonde sur cette liste de mots. Là encore, les résultats quantitatifs suggèrent des différences entre les sous-corpus, même si elles peuvent sembler faibles. On aurait, d'après ces proportions différentes, une sorte de continuum de la simplification lexicale, avec un des sous-corpus contenant deux fois moins d'items lexicaux "complexes" que les deux autres (EN-CA-Résumés), le corpus EN-Brochures étant le plus complexe des trois, et le corpus EN-CAB, se situant entre les deux.

Il convient maintenant de comparer la proportion de ces items pour chaque sous-corpus PLAIN par rapport au sous-corpus LEX (*expert-to-expert*) dont il propose une reformulation. Pour rappel, les corpus EN-Brochures et EN-CAB s'appuient et recontextualisent le discours législatif ; ils sont donc chacun comparés au corpus EN-Lois. Le sous corpus EN-CA-Résumés est quant à lui comparé au sous-corpus EN-CA-Jugements.

Les sous-corpus EN-Brochures et EN-CAB contiennent respectivement deux et trois fois moins d'items lexicaux non-approuvés que le sous-corpus EN-Lois. Ainsi, si le corpus EN-Brochures présente une plus grande complexité lexicale, il y a cependant une diminution de la densité de mots à éviter dans le passage des textes législatifs aux brochures, d'un facteur 2. EN-CAB présente une proportion encore plus faible de ces éléments lexicaux. Or, le sous-corpus EN-Brochures est issu d'une plus grande variété d'institutions, ce qui expliquerait peut-être cette différence. Comme nous l'avons noté pour les scores de lisibilité, une plus grande diversité des sources discursives pourrait ainsi donner lieu à une plus grande variabilité dans la fréquence d'emploi de ces items lexicaux. Concernant le discours juridictionnel et sa reformulation, on voit également une forte diminution de la fréquence d'items lexicaux "interdits" dans le passage du corpus EN-CA-Jugements au corpus EN-CA-Résumés, ce dernier en contenant trois fois moins que le premier. Tous les genres formant le corpus EN-PLAIN suggèrent donc une importante simplification lexicale dans le passage d'un lectorat-cible à l'autre, sans qu'ils adhèrent pour autant totalement aux recommandations.

On peut remarquer que la différence dans la proportion d'éléments lexicaux jugés complexes entre les deux genres du corpus EN-LEX se retrouve entre les genres discursifs du corpus EN-PLAIN. Cela pourrait suggérer que la plus grande complexité lexicale du discours législatif par rapport au discours judiciaire se retrouve dans une comparaison entre les textes procéduraux et les résumés de jugements. Bien sûr, il est aussi possible que les textes du discours judiciaire contiennent tout autant de mots lexicaux complexes, qui ne sont pas capturés par cette liste.

7.1.2 Exploration des changements lexicaux par sous-corpus et discussion

Nous avons jusqu'ici dégagé des tendances globales concernant la proportion des mots non-approuvés dans les différents sous-corpus. Nous proposons maintenant d'examiner en détail les items lexicaux non-approuvés les plus fréquents, et leurs fréquences dans le passage du discours juridique au discours de reformulation. Dans la suite de ce chapitre, la fréquence normalisée des termes est donnée par million de mots (pmm).

Fréquence du vocabulaire non-approuvé dans le discours législatif et sa reformulation

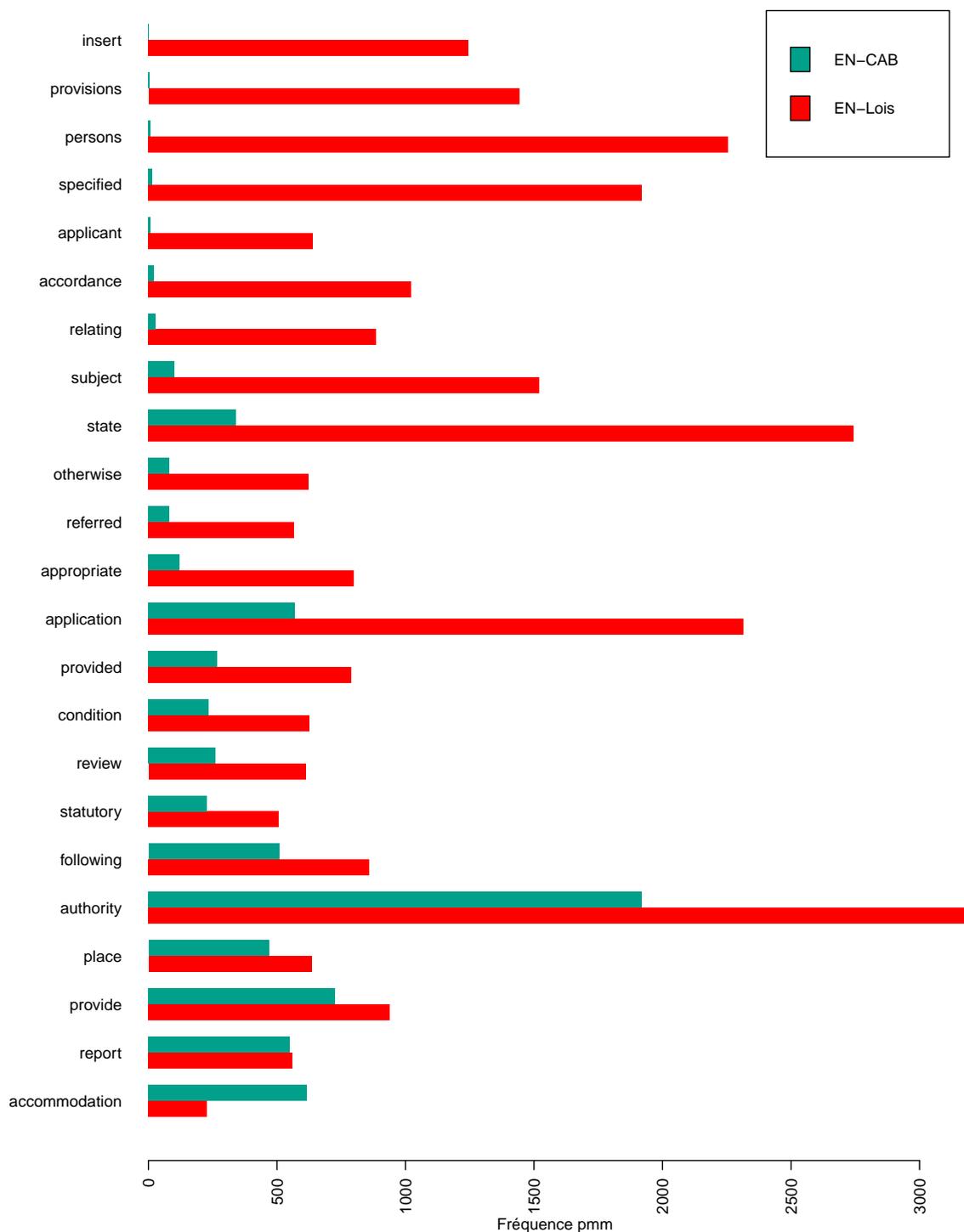
Les figures 7.1 et 7.2 montrent les items lexicaux non-approuvés les plus fréquents (apparaissant au moins 500 fois pmm) dans le sous-corpus expert-à-expert EN-Lois (en rouge sur les figures), et dans les sous-corpus EN-Brochures et EN-CAB (en vert sur les figures).

Concernant le lexique non-approuvé dans EN-Lois, les mots très fréquents sont typiquement associés au genre discursif du texte de loi. Les mots comme *insert*, *amendment provision* sont typiques de la dimension méta-linguistique de ces textes, puisqu'ils permettent en général d'indiquer les amendements des textes de lois précédents, de renvoyer à d'autres textes législatifs passés ou futurs, ou de faire référence au texte lui-même, comme on le voit dans les exemples 77 et 78.

- (77) This is a reprint of the Crimes Act 1961 that incorporates all the **amendments** to that Act as at the date of the last **amendment** to it. (EN-Lois)
- (78) After subsection **insert**– "This section does not apply to a consumer contract (but see the **provision** made about such contracts in section 31 of the Consumer Rights **Act** 2015)." (EN-Lois)

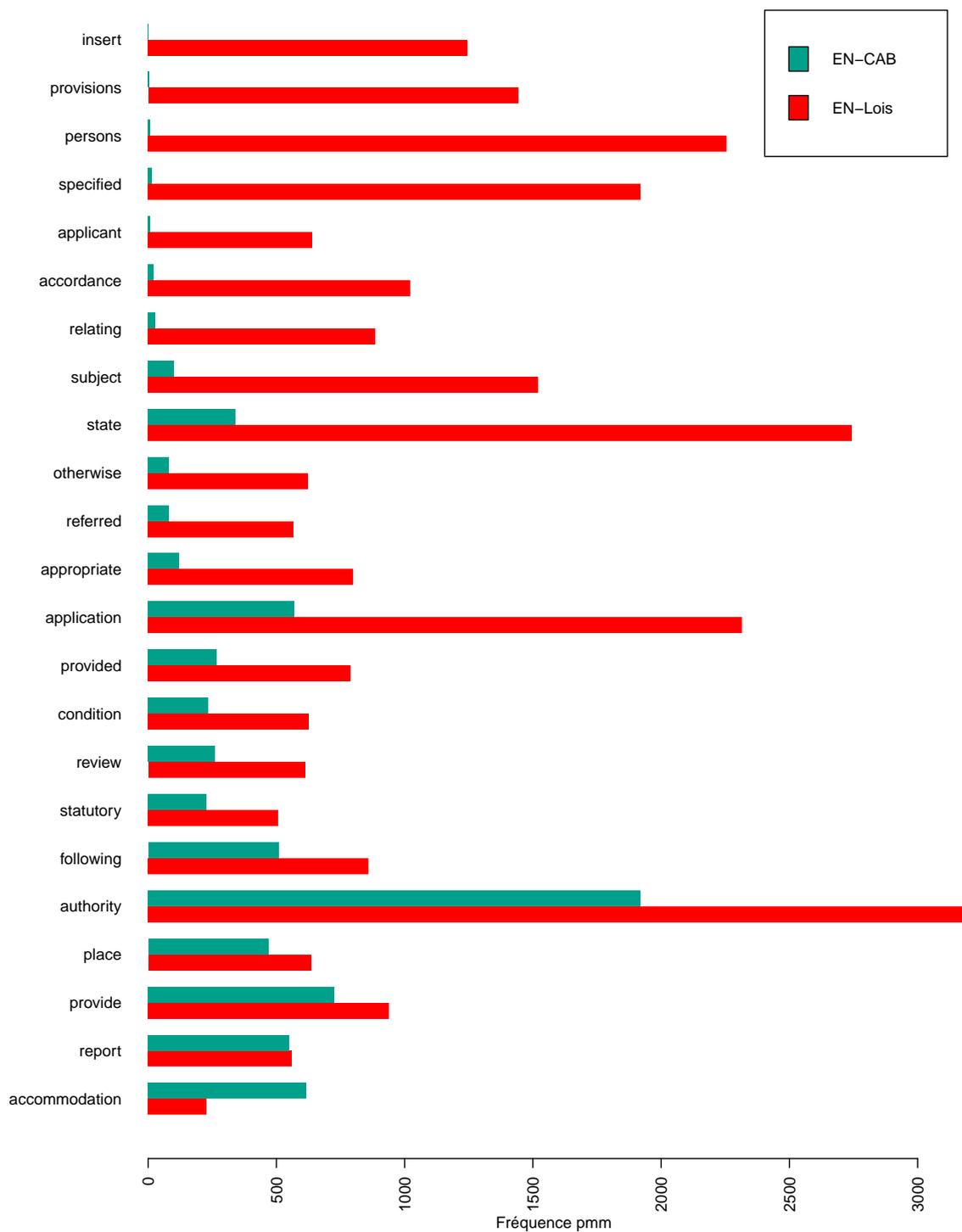
Certains mots apparaissant dans le corpus EN-Lois font partie de séquences phraséologiques typiques du discours juridique, en particulier législatif, comme par exemple *relating*, qui a une fréquence de 687,5 pmm et apparaît dans *relating to* (ex : *settlements relating to infringements of competition*), ou *subject to*, d'une fréquence de 1095,8, qui a des emplois tels que *This section is subject to section 52 (parental etc. consent)* (EN-Lois). Il n'est pas surprenant que ces éléments lexicaux et phraséologiques, très spécifiques au genre discursif spécialisé, soient très fréquents dans EN-LEX, mais pas dans les textes destinés aux justiciables. Appartenant à un registre soutenu, ils sont donc bien évités dans les textes reformulant le discours législatif. Néanmoins, certains éléments lexicaux, qui semblent particulièrement liés à la phraséologie juridique apparaissent, bien que discrètement, dans les deux sous-corpus EN-Brochures et EN-CAB. *Accordance*, par exemple, a une fréquence 1000 pmm dans le corpus EN-Lois, et apparaît sous la forme d'une préposition complexe (d'après le terme utilisé par Bhatia (1994)) très spécifique à la phraséologie du droit et en particulier au discours législatif (comme dans *in accordance with the framework determined by the Government*). Or, cette expression n'est pas totalement absente des sous-corpus dans EN-Brochures (67,8 pmm) et EN-CAB (18,8 pmm). Même si

FIGURE 7.1 – Proportion des mots non-approuvés les plus fréquents dans EN-Lois et EN-Brochures



ces fréquences peuvent sembler faibles, il est étonnant qu'on retrouve cet élément si spécifique au discours législatif dans les sous-corpus expert-à-non-expert. Nous

FIGURE 7.2 – Proportion des mots non-approuvés EN-Lois et EN-CAB



reproduisons ci-dessous deux passages qui contiennent la préposition complexe *in accordance with* :

(79) Contact the sender directly If you want to stop getting mail from a particular sender, contact them directly. Include : your full name and address the date this sentence : "Please stop processing my personal data for direct marketing purposes **in accordance with** the Data Protection Act 1998." (EN-CAB)

(80) "You have the right to be treated with dignity and respect, **in accordance with** your human rights." (Section 3a of the NHS Constitution) (EN-Brochures)

The authority can confirm, reverse, or amend a decision **in accordance with** the Customs and Excise Act 1996. (EN-Brochures)

Dans ces extraits, l'expression "*in accordance with*" est soit une citation directe du texte de juridique de référence (la *NHS Constitution* par exemple) ; soit elle est employée dans un exemple de courrier administratif à envoyer par le justiciable. D'une part, il arrive donc aux rédacteur.ices de transmettre des connaissances juridiques sur les droits des citoyen.es en citant directement les textes de loi(exemple (79)). D'autre part, des passages procéduraux sont également incorporés au discours explicatif afin de présenter les démarches à suivre mais aussi la phraséologie à adopter par les justiciables pour s'adresser aux acteurs juridiques et administratifs. Dans l'exemple de EN-CAB (78), les rédacteur.ices expliquent ainsi qu'il faut citer le texte de loi en utilisant la préposition complexe [*in accordance with* + Nom de la loi] dans les courriers administratifs pour demander la fin du traitement de leurs données personnelles. Ces textes incorporent donc la phraséologie juridique par endroits, par exemple pour indiquer aux justiciables comment communiquer avec les instances administratives. Ces exemples soulignent le caractère hybride du discours de diffusion du droit. Preite (2018) décrit ce discours rapporté comme des "éléments de dialogisme interdiscursif montré", c'est-à-dire le signalement explicite du discours autre (législatif) avec lequel le texte entre en interaction par le biais de la médiation.

Certains mots non-recommandés pour rédiger en style clair sont à la fois très fréquents dans EN-Lois et dans les sous-corpus destinés aux justiciables, comme on le voit dans les figures 7.1 et 7.2. C'est le cas en particulier de *authority*, qui a une fréquence de 3058 pmm dans EN-Lois, mais également de 1 496,8 pmm dans EN-Brochures et de 2028,6 dans EN-CAB. Il s'agit du mot de la liste le plus fréquent pour EN-Lois et EN-CAB, et l'un des plus fréquents pour EN-Brochures.

(81) The director of public health for a local **authority** must prepare an annual report on the health of the people in the area of the local authority. (EN-Lois)

(82) Each local **authority** has a director of public health who is responsible in law for exercising public health functions of the local **authority**, and publishes an annual report stating what progress has been made towards improving the local population's public health. (EN-Brochures)

- (83) The local **authority** can offer different types of assistance to improve living conditions, including : grants loans assistance from home improvement agencies supporting purchase and relocation other forms of assistance. (EN-CAB)

Les exemples ci-dessus illustrent l'emploi de ce mot dans la collocation *local authority*. Cette collocation est utilisée Dans tous les extraits présentés, spécialisés ou s'adressant aux justiciables. Cependant, lorsqu'on regarde la liste d'alternatives proposées par PLC, l'alternative suggérée est la suivante : *right, power, may (as in 'have the authority to')*. La règle sur l'emploi de *authority* porte sur une construction à verbe support, non sur la collocation *local authority*. On trouve quelques occurrences de cette expression dans les deux sous-corpus pour expliquer certains droits et pouvoirs des justiciables et institutions : 4,2 pmm dans EN-Brochures (*The IPs may : not have the authority to make decisions about their child's education and medical care*) et 3,8 dans EN-CAB (*The CMS has the authority to trace the parent who should pay maintenance*). On trouve également la variante *have authority to*. Ces exemples confirment que les recommandations portent non seulement sur des éléments lexicaux isolés, mais aussi la phraséologie.

On notera que certains mots non-approuvés sont morphologiquement ambigus, comme *place*, qui est présent dans la liste en tant que verbe, et qui doit être selon ses rédacteur.ices remplacé par le verbe plus fréquent *put*. On trouve quelques occurrences du verbe *placed* dans EN-Lois et même dans EN-Brochures.

- (84) If it is not practicable or appropriate to **place** the child or young person with any of the persons specified in any of paragraphs to of subsection , the chief executive may place the child or young person in a residence. (EN-Lois)
- (85) The Home Office will look at the following factors in deciding if it should **place** you in detention. (EN-Brochures)

Cependant, la plupart des occurrences dans les trois sous-corpus correspond au nom *place*. C'est le cas également pour *state*, qui est un verbe dans la liste d'alternatives, mais apparaît en général dans un groupe nominal (en particulier *Secretary of State* dans EN-Lois). Le mot *application* est également de ceux-là, puisque les recommandations pour un style clair préconisent d'utiliser plutôt son alternative *use, or*, il est l'un des mots non-approuvés les plus fréquents dans EN-Brochures et EN-CAB. Cependant, il est employé dans ces sous-corpus, pour référer non pas à l'application d'un concept juridique à une situation, mais à une procédure administrative de demande d'aides sociales, de visa ou autre démarche administrative, ou au document qui en résulte, comme on le voit dans l'exemple ci-dessous.

- (86) If you make your **application** after your visa has ended, you'll usually lose the right to work, access education and receive benefits while you wait for a new visa. (EN-CAB)

Les résultats quantitatifs doivent donc être considérés avec précaution, pour certains mots qui sont morphologiquement ambigus. D'autres mots non-recommandés sont très fréquents dans les sous-corpus EN-CAB et EN-Brochures sans pour autant être ambigus, et sont, peut-être de manière surprenante, plus fréquents dans ces deux sous-corpus que dans le sous-corpus destinés aux expert.es. C'est le cas par exemple du nom *accommodation*, qui d'après PLC devrait être remplacé par son alternative claire et simple *home* ou *where you live*. *Accommodation* est plus fréquent dans les corpus EN-Brochures (652,9 pmm) et EN-CAB (561,2 pmm) que dans EN-Lois (195,8 pmm).

- (87) When you claim asylum we will ask you if you need **accommodation**. If you have nowhere to stay, or insufficient funds to provide for your own **accommodation**, we will firstly accommodate you in what is called "**initial accommodation**" (IA). (EN-Brochures)
- (88) You'll be trespassing if you stay in the **accommodation** without your landlord's permission either after the fixed term has ended or after the notice period has expired. (EN-CAB)

Dans le premier exemple, on voit que ce mot est employé comme un terme spécifique à la procédure administrative de demande d'asile, pour définir un type particulier de logement (*initial accomodation*), mais *accomodation* en lui-même n'est pas défini. Dans l'extrait d'EN-CAB, *accomodation* ne fait pas non plus l'objet d'une définition ou d'une reformulation. Les rédacteur.ices de ces textes supposent peut-être que le lectorat connaît déjà ce synonyme plus formel de *home*, puisque le choix de lire ce texte émane probablement d'une démarche administrative en cours et que le lecteur a donc déjà été exposé à ce mot. Dans le corpus EN-Brochures, trois verbes en particulier apparaissent davantage que dans le corpus spécialisé EN-Lois : *provide*, *ensure*, *report*. *Provide*, selon la liste de PLC, devrait être remplacé par son alternative claire et simple *give*, *ensure* par *make sure* et *report* par *tell*.

- (89) Remember, by acting as a witness in a criminal case, you are helping **ensure** a person is given a fair trial. (EN-Brochures)
- (90) You will often be asked to **provide** personal information to various agencies and organisations. (EN-Brochures)

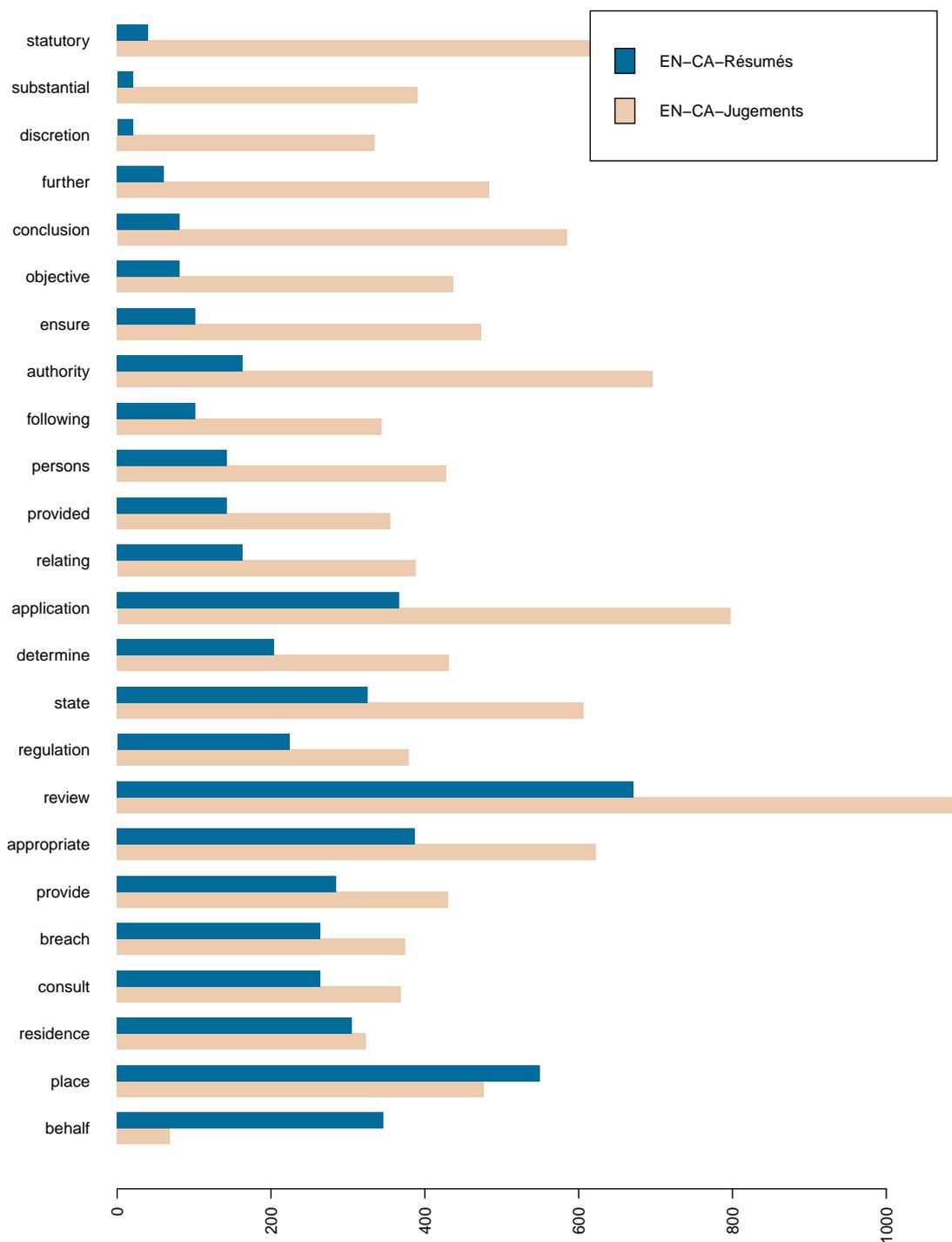
(91) **Report** [the crime] to the police by calling 111 or going to the Police Station. (EN-Brochures)

La relativement haute fréquence de *ensure* dans EN-Brochures peut s'expliquer par le fait que ce verbe est proche de l'alternative claire et simple suggérée *make sure*, puisqu'on y retrouve la racine *sure*, et que son sens pourrait être facilement déduit par les lecteur.ices en se fondant sur sa morphologie ; il serait donc possible de prédire son sens en se fondant sur leurs propres connaissances (inconsciente) de la langue, comme le suggèrent Bailin et Grafstein (2001). Concernant *provide* et *report*, ces verbes sont liés à des actions et à des formulations typiques dans le cadre de procédures judiciaires et administratives. Comme pour *accommodation* plus haut, on peut supposer que les rédacteur.ices supposent que ces items lexicaux font partie des connaissances partagées entre le lectorat, l'institution et eux-mêmes, c'est-à-dire que les personnes ont déjà rencontré ce mot. Pour *report*, en particulier, on peut supposer que le lectorat a déjà rencontré ce mot par ailleurs, dans des séries policières ou dans les médias. Une autre explication possible est que les rédacteur.ices adoptent le lexique même des institutions pour illustrer leur style, et faciliter la communication avec elles. Ainsi, si une personne se rend dans un commissariat pour signaler un délit ou pour porter plainte, c'est le mot *report* et non *tell* qui sera utilisé par ses interlocuteur.ices. Il faudra qu'elle-même utilise pour indiquer la raison de sa présence (*to report a crime*). On peut donc supposer que les rédacteur.ices utilisent le lexique de l'administration, de la police ou du tribunal pour aider les justiciables non seulement à comprendre mais aussi à interagir avec les institutions judiciaires et administratives. Nous faisons l'hypothèse que l'utilisation de certains de ces éléments lexicaux considérés par PLC comme complexes serait donc liée à la fonction procédurale de ces textes, afin d'accompagner les justiciables dans ces procédures, afin de les préparer à ou de leur confirmer le parler spécifique des institutions.

Vocabulaire non-approuvé dans EN-CA-Résumés vs. EN-Jugements

Nous examinons maintenant les éléments lexicaux non-approuvés dans les motifs de la CS du Canada et leurs résumés. Ces contiennent une plus faible proportion des items lexicaux étudiés, et sont donc représentés dans la Figure 7.3 les items lexicaux apparaissant plus de 300 pmm dans l'un ou l'autre des sous-corpus. Certains éléments lexicaux non-approuvés sont particulièrement saillants dans les jugements, en particulier *statutory* (782,2 pmm) et *conclusion* (600,8 pmm), dont la fréquence diminue fortement dans EN-CA-Résumés (elle passe à 90,8 pmm pour ces deux

FIGURE 7.3 – Proportion des mots non-approuvés les plus fréquents dans EN-CA-Jugements et EN-CA-Résumés



mots). Ces éléments lexicaux sont liés à des concepts centraux au processus de contrôle judiciaire dans EN-CA-Jugements :

- (92) The extent to which the **statutory** powers under the Draft Federal Act permit the regulation of these matters remains circumscribed by the requirement of systemic risk (...).
- (93) The question is whether the tools of **statutory** interpretation – including the text, context and purpose of the provision – can reasonably support the Five judges came to the same **conclusion**, but for different reasons. (...). (EN-CA-Jugements)
- (94) I base my **conclusion** on six interrelated considerations, which I will examine in turn. (EN-CA-Jugements)

Statutory apparaît dans des collocations telles que *statutory powers* et *statutory interpretation* qui sont liées au rôle fondamental d'interprétation par l'institution elle-même, plus haute juridiction canadienne, et aux références fréquentes aux pouvoirs législatifs. Le mot *conclusion* renvoie quant à lui à la fois aux décisions des juges de la CSC eux-mêmes et aux commentaires sur les décisions des cours inférieures. Ces éléments ne sont pas entièrement absents de EN-CA-Résumés, mais ils y sont rares. *Conclusion* n'apparaît que 5 fois dans ce sous-corpus, en particulier pour introduire les différences entre les motifs des juges de la CSC, lorsque la décision n'est pas rendue à l'unanimité, comme on le voit dans l'exemple (95). Le corpus de résumés contient par ailleurs seulement quatre occurrences de *statutory* :

- (95) Five judges came to the same **conclusion**, but for different reasons. (EN-CA-Résumés)
- (96) Both common law (made by courts) and **statutory** law (made by Parliament) are equally valid. But Parliament can change the common law by passing a statute (like the Act), if it wants to. Where **statutory** law (like the Act) doesn't cover something, the common law can fill in the blanks. (EN-CA-Résumés)

L'exemple (96) montre que *statutory* est utilisé ici pour rappeler les grands principes du système de *common law*, alors que dans les jugements cités ci-dessus, il est employé dans des collocations que l'on peut qualifier de "plus spécialisées". Dans EN-CA-Résumés, il semble utilisé pour donner des informations générales sur le fonctionnement du droit et les interactions entre pouvoirs législatifs et pouvoirs judiciaires.

Plusieurs des éléments lexicaux apparaissant comme très fréquents dans les jugements d'après la Figure 7.3 semblent liés à la nature argumentative des textes que constituent les motifs de jugement. C'est le cas de *conclusion* que nous venons de voir, mais aussi de certains autres éléments lexicaux, en gras ci-dessous :

- (97) The first step is to identify the **objective** of the impugned changes and, in turn, to **determine** whether it is pressing and **substantial**. (EN-CA-Jugements)

Dans cet extrait, *substantial* (également *pressing*, mais qui n'est pas dans notre liste) est en fait un mot utilisé dans le texte de loi sur lequel le juge s'appuie, rappelant l'importance des éléments lexicaux génériques et vagues pour les juges : ici, c'est bien cet item lexical qui est au cœur de l'argumentaire et de ce qu'il convient de démontrer. Les mots *objective*, *determine*, quant à eux, servent à la juge² à exposer les première étapes de sa démonstration. On peut d'ailleurs remarquer la proximité entre ces extraits de discours judiciaires et un texte du discours académique qui décrirait les objectifs d'une étude ; or, les éléments lexicaux *determine*, *conclusion* mais également *following*, *further* font partie à la fois du vocabulaire fréquent dans les jugements et dans la liste des mots-clés du vocabulaire académique de Gardner et Davies (2014). Il n'est donc pas étonnant que ces items lexicaux soient fréquents dans les jugements, mais n'apparaissent que peu dans les résumés, témoins d'un effort de simplification lexicales.

Certains éléments lexicaux non-approuvés sont cependant fréquents à la fois dans les jugements et leurs résumés, en particulier *review*, dont l'alternative claire et simple suggérée est *look at again*. Cependant ce mot est également morphologiquement ambigu. Nous avons inclus le verbe *review* comme élément lexical complexe, mais sous sa forme nominale, il s'agit d'un terme qui est au cœur du système de *common law* ; cet exemple souligne les frontières poreuses entre ce que nous avons considéré comme appartenant à un lexique complexe et la terminologie.. Dans le sous-corpus EN-CA-Jugements, *review* est souvent employé comme verbe (417,0 pmm pour le lemme verbal) mais la plupart des occurrences correspondent au nom *review* (938,9 pmm pour le lemme nominal), comme on le voit dans ces deux extraits :

- (98) When courts **review** administrative decisions for compliance with the Charter, Charter rights must be the focus of the inquiry (...)

(99) I agree that the **standard of review** is reasonableness.

- (100) Based on my **review** of the jurisprudence, Mr. Suter's facts are, indeed, entirely unique.

Il apparaît notamment dans le terme *standard of review*, qui fait référence à la norme de contrôle que les juges peuvent appliquer dans leurs arrêts (en droit administratif), lorsqu'ils considèrent une décision des cours inférieures. Le mot *review* est

2. L'extrait est tiré de motifs de jugements rendus par la juge Rosalie Abella.

également le mot non-approuvé le plus fréquent de la liste dans EN-CA-Résumés. Au vu de la fonction de la CS, il semble logique que le verbe *review* ne puisse pas être évité, puisque l'activité fondamentale de la Cour suprême est la révision judiciaire (*judicial review*). On le trouve là aussi à la fois en tant que verbe et que nom, pour rappeler le parcours juridictionnel des cas dont il est question, ainsi que l'activité centrale de la CS :

(101) Mr. Wall asked the Alberta Court of Queen's Bench to **review** the decision.

(102) This case involved courts' **review** of an administrative tribunal's decision. (EN-CA Résumés)

Ces exemples montrent que les résumés n'utilisent pas toujours ce terme sous forme verbale mais également sous forme nominale, comme on le voit en comparant les exemples (101) et (102), ce dernier contenant le mot sous la forme d'un GN complexe (*review + of + GN*), forme . On peut de nouveau faire l'hypothèse que les rédacteur.ices considèrent que ces éléments lexicaux ou phraséologiques doivent être utilisés pour pouvoir expliquer le plus précisément et correctement possible le fonctionnement et le rôle de la CS. On trouve dans le corpus des exemples d'adoption de la phraséologie spécialisée par l'utilisation de mots non-approuvés dans des collocations spécifiques :

(103) This decision noted that **breaching** someone's language rights can be seen as a sign of bias against a language group.

Dans cet exemple, le verbe *breach* dans la collocation *breach someone's rights* ne pourrait être ici remplacé par son alternative suggérée, qui est *break*, puisque *?breaching someone's rights* ne forme pas une collocation. Là encore, on voit que les textes de diffusion du discours juridique adoptent les collocations et la phraséologie spécialisées, car l'alternative claire et simple ne peut pas toujours exprimer le concept juridique dont il est question.

Enfin, comme pour la diffusion du discours législatif décrite plus haut, certains mots jugés complexes dans les recommandations apparaissent davantage dans les textes destinés au grand public, comme *place*, d'une fréquence de 19,3 pmm dans EN-Jugements contre 567,5 pmm dans EN-CA-Résumés. Comparons :

- (104) Charter rights (...) do not **place** the government under an obligation to act in order to obtain specific societal results such as the total and definitive eradication of genderbased pay inequities in private sector enterprises. (EN-CA-Jugements)
- (105) He said he **placed** Ms. Jordan's ashes near her family cottage, because that's what Ms. Jordan had said she wanted done with her ashes if she died. (EN-CA-Résumés)
- (106) The appeal was based on a pre-trial motion and the courts have not decided whether copyright infringement actually took **place**. (EN-CA-Résumés)

Dans son emploi verbal, c'est-à-dire les deux premiers exemples, il est évident que *place* est utilisé dans un sens beaucoup plus concret, pour rappeler les faits de l'affaire jugée dans EN-CA-Résumés, que dans l'expression verbo-nominale abstraite *place someone under the obligation to* dans le jugement. On peut supposer qu'il est considéré comme facilement compréhensible par le public dans le contexte (*near her family cottage*), et qu'il n'est pas nécessaire d'employer son alternative *put*. Surtout, si ce mot est si fréquent dans le corpus EN-CA-Résumés, c'est qu'il y est employé (comme dans EN-CAB et EN-Brochures) en tant que nom, ou dans la construction figée *take place* comme le montre l'exemple (106).

En revanche, *behalf*, qui a une fréquence de 340,5 pmm contre 60,5 pour EN-Jugements n'est pas ambigu, et apparaît dans le groupe prépositionnel complexe *on behalf of*, auquel il est pourtant recommandé de substituer *for* :

- (107) The Attorney General, **on behalf of** the federal government, agreed that the rule breached section 3.

Selon les rédacteur.ices, cette expressions serait peut-être connue du lectorat, ou compréhensible dans ce contexte. Globalement, ces résultats suggèrent que les recommandations lexicales ne sont pas entièrement respectées, même si les résultats doivent être abordés avec précaution au vu d'ambiguïtés morphologiques, et qu'ils diffèrent selon les sous-corpus. Plusieurs hypothèses peuvent expliquer la présence de ces éléments lexicaux jugés complexes. D'une part, il faut rappeler que la liste a été constituée par les membres d'un organisme promouvant le style clair. Il y a donc une certaine part de subjectivité dans les éléments lexicaux choisis : certains mots qui semblent complexes pour les éditeur.ices de l'association PLC ne le sont peut-être pas pour les rédacteur.ices des textes du corpus PLAIN. D'autre part, comme nous l'avons vu dans les exemples, certains mots paraissent difficiles à éviter dans les contextes de production des textes et sont nécessaires dans les textes de médiation juridique. Ils font partie intégrante du style bureaucratique et juridique

qu'il est nécessaire d'employer pour expliquer aux justiciables leurs droits, mais aussi les procédures judiciaires, administratives et le rôle des institutions. Parfois, ces éléments lexicaux font l'objet de stratégies d'explicitation, comme nous le verrons après avoir examiné la densité terminologique.

7.2 La densité terminologique dans les corpus EN-LEX et EN-PLAIN, une mesure du degré de spécialisation ?

La fréquence des termes juridiques étudiés change-t-elle dans le passage du corpus EN-LEX à EN-PLAIN ? C'est la question à laquelle nous allons nous efforcer de répondre, en nous fondant sur la liste de termes établie à partir du glossaire présenté dans l'Annexe C.2.

7.2.1 Proportion de termes juridiques : résultats quantitatifs

Les résultats quantitatifs sont présentés dans le Tableau 7.3 ci-dessous. Nous rappelons que la densité terminologique est définie ici comme la proportion de termes simples (c'est-à-dire un terme constitué d'un seul mot), l'outil utilisé ne prenant pas en charge les termes complexes.

TABLEAU 7.3 – Proportion de termes dans les différents sous-corpus

Corpus	Proportion de termes (%)
EN-Lois	5,1
EN-CA-Jugements	5,3
EN-CA-Résumés	5,9
EN-Brochures	4,1
EN-CAB	3,3

Les résultats semblent moins contrastés que dans la section précédente, où la diminution de la proportion d'items lexicaux de la liste entre EN-LEX et EN-PLAIN apparaît très nettement. On voit tout d'abord que la densité de termes par rapport au nombre total de mots est similaire dans les deux sous-corpus spécialisés EN-Lois et EN-CA-Jugements, puisque chacun contient environ 5% de termes de la liste. Les deux sous-corpus présentant le même degré de spécialisation, il n'est pas étonnant qu'ils contiennent la même proportion de termes. Plus surprenants sont les résultats

concernant la densité terminologique dans les sous-corpus de EN-PLAIN, puisque la densité terminologique ne diminue pas. Dans EN-CA-Résumés, la proportion de termes est même très légèrement supérieure à celle du corpus spécialisés correspondant, EN-CA-Jugements (5,9% contre 5,3%). Cette proportion moyenne des termes, qui est sensiblement la même voire plus haute plus haute dans les textes destinés aux non-spécialistes que dans ceux destinés aux spécialistes s'explique peut-être par le genre discursif représenté par EN-CA-Résumés : il s'agit de résumés, des textes courts, où quelques occurrences de termes peuvent avoir une incidence importante sur la densité moyenne de ces termes par rapport à tous les mots du texte. Concernant les textes des sous-corpus EN-Brochures et EN-CAB, qui simplifient le discours législatif, la densité des termes diminue respectivement d'environ 20% et un tiers par rapport à la densité terminologique dans EN-Lois, ce qui suggère que la densité terminologique, bien qu'elle diminue, reste relativement haute dans ces textes également. Notre question de recherche était de savoir si la densité terminologique peut servir de mesure du degré de spécialisation des textes en permettant de distinguer quantitativement les textes destinés aux spécialistes et ceux destinés aux non-spécialistes. Devant ces résultats, force est de constater que la densité terminologique ne semble pas refléter la différence entre les corpus EN-LEX et EN-PLAIN aussi clairement que les résultats de Ferraresi (2019). Cependant, comme dans ses résultats, la mesure de la densité terminologique suggère que les sous-corpus de diffusion du droit d'EN-PLAIN ne forment pas un bloc homogène, et que les genres discursifs présentent des variations subtiles (Ferraresi 2019). Ainsi le sous-corpus EN-CAB est celui qui contient le moins de termes de la liste, et EN-CA-Résumés celui qui en contient le plus par rapport au nombre total de mots.

Comment expliquer que la densité terminologique ne diminue pas autant que le lexique considéré complexe dans le passage de EN-LEX et EN-PLAIN ? D'une part, il convient de rappeler que notre liste *ad hoc* ne contient qu'un nombre restreint de termes mono-lexicaux, et que certains termes mono- ou poly-lexicaux voient peut-être leur proportion diminuer fortement lors du passage de l'un à l'autre registre. D'autre part, on peut faire l'hypothèse que c'est ces termes spécialisés sont nécessaires pour expliquer les procédures et leurs droits aux justiciables. Il paraît en effet difficile de parler de certaines procédures et du fonctionnement du système judiciaire, sans utiliser les termes renvoyant aux différentes juridictions et aux documents, étapes et tribunaux ou instances administratives. Ces résultats quantitatifs, bien qu'ils soient incomplets, sont intéressants car ils suggèrent que ce n'est pas la proportion de termes spécialisés qui change le plus entre le discours du droit et les textes à destination du grand public, mais la proportion du lexique considéré comme

difficiles par PEC, car ils appartiennent à un registre soutenu ou académique, et ne correspondent pas au vocabulaire commun ou "familier" (toujours pour un lectorat-modèle). Il convient maintenant d'observer plus en détail quels termes sont les plus fréquents dans les deux corpus en anglais.

7.2.2 Termes les plus fréquents dans les corpus EN-LEX et EN-PLAIN

Quels sont les termes du glossaire les plus fréquents dans les différents sous-corpus expert-à-expert et ceux destinés au grand public ? Les Tableaux 7.4 et 7.5 présentent la fréquence des termes du glossaire dans les discours législatif et judiciaire avant et après simplification.

TABLEAU 7.4 – Termes de la liste les plus fréquents dans le discours législatif avant et après simplification

(a)		(b)		
Terme	Fréquence dans EN-Lois (pmm)	Terme	Fréquence dans EN-CAB (pmm)	Dans EN-Brochures (pmm)
act	12995	child	2680,8	6688,8
order	5523	court	2630,1	2812,2
court	3256	support	1028,7	3046,2
child	2663	act	1109,6	1752,5
service	1826	service	1717,8	1513,9
case	1514	order	2209,5	1357,9
notice	1172	case	1004,1	1559,8
adoption	1072			

Que remarque-t-on globalement concernant les termes les plus fréquents dans les différents corpus ? Si l'on adopte une vision de la terminologie comme gradable, c'est-à-dire où des termes présentent degré de spécialisation plus ou moins haut, on remarque que les termes qui sont les plus fréquents semblent avoir un faible degré de spécialisation, dans la mesure où ils appartiennent aux termes-clés du droit ou *core termes* selon l'expression de Ferraresi (2019). Cornu définit ces termes comme des termes polysémiques d'appartenance juridique principale, ayant un sens juridique principal, fort, et ayant pris un sens dérivé dans des usages généraux de la langue (Cornu 2005 : 69). Ces termes peuvent être classés en s'inspirant de la classification de Cornu, en différentes catégories. Cornu distingue parmi ces termes "porteurs des notions fondamentales du droit" (Cornu 2005 : 71) : les termes-clés

TABLEAU 7.5 – Termes de la liste les plus fréquents dans le discours judiciaire anglophone avant et après simplification

(a)		(b)	
Terme	Fréquence dans EN-CA - Jugements (pmm)	Terme	Fréquence dans EN-CA-Résumés (pmm)
court	4 416,7	court	10136,2
law	4 233,1	law	4884,9
appeal	3 409,8	judge	4498,2
act	2 748,2	case	4131,9
judge	2 478,6	trial	3215,9
case	2 283,4	appeal	3093,8
trial	1 675,7	justice	2686,7
charter	1 651,0		

du vocabulaire de base, incluant *act*, *law* et *justice* ; les termes-clés du vocabulaire judiciaire, dont font partie *judge*, *case*, *trial*, *appeal* et *court* ; les termes-clés des opérations juridiques principales et des actes (= documents) juridiques courants, c'est-à-dire des documents comme *notice* ou *order* ; et un lot divers d'autres termes-clés, comme *divorce*. Certains de ces termes-clés sont plus spécifiques à un sous-registre du droit en particulier. Par exemple, le terme le plus fréquent dans le corpus EN-Lois est *act*, puisque les textes qui le composent font constamment référence, au sein du texte, au texte de loi lui-même, mais aussi à d'autres lois affectées ou affectant cette même loi, comme on le voit dans la phrase ci-dessous, qui contient pas moins de quatre occurrences du terme *act* :

- (108) In this section, "enactment" includes— an enactment contained in subordinate legislation (within the meaning of the Interpretation **Act** 1978), and an enactment contained in, or in an instrument made under, an **Act** of the Scottish Parliament, an **Act** or Measure of the National Assembly for Wales or Northern Ireland legislation, and references to an enactment include a reference to an enactment passed or made after the passing of this **Act**.

On note dans ce passage l'emploi récurrent de *enactement* (non présent dans la liste de termes), ce qui rappelle à nouveau que notre liste de termes est incomplète. Dans EN-Lois, on trouve aussi des termes liés aux actes juridiques principaux :

- (109) The holder of a licence may at any time surrender that licence by written **notice** to the Registrar.

- (110) If a party has given **notice** under subsection and it is no longer possible or desirable for the witness to give evidence by the means stated in the **notice**, the party may file an amended **notice** but must do so as early as practicable before the case is to be tried.
- (111) If the court determines a question under subsection , the court may make an **order**– varying or terminating the authorisation (...) (EN-Lois)

Il s'agit de termes qui font références à des documents juridiques publiés par les tribunaux et autres acteurs administratifs. On note par ailleurs que ces termes sont insérés dans des prédicats complexes, *make an order* et *give (someone) notice*, dont nous reparlerons plus loin. Le discours judiciaire représenté par EN-CA-Jugements, bien que contenant également de nombreuses occurrences du terme *act* sous forme de références et de renvois à la loi, est bien davantage caractérisé par des termes-clés de la terminologie judiciaire, en particulier *court*, *law* et *appeal*. Nous prenons ici l'exemple de *law*.

- (112) It follows that a claimant alleging that a **law** violates [Section] 121 must establish that the **law** in essence and purpose restricts trade across a provincial border. (EN-CA-Jugements)
- (113) This Court has previously recognized the risk of jurisdictional overreach in these types of cases (...) : At **common law**, the tort of defamation crystallizes upon publication of the libellous material This also raises difficult issues when publication occurs through the Internet (Banro, at para 3) (EN-CA-Jugements)

L'exemple (112) fait référence au concept de loi en tant que texte normatif que peut contester un.e demandeur.esses, mais certaines occurrences de *law* dans EN-CA-Jugements correspondent en réalité à des termes complexes sous forme de noms composés (*compounds*) et de GN complexes, tels que *mistake of law* ou des hyperonymes de *law*, par exemple *defamation law*. Dans l'exemple (113), le terme correspond non pas au terme français *loi* mais en fait au terme complexe désignant le système juridique de *common law* caractéristique du droit anglo-saxon. Cet extrait contient également le terme *Court* ; le ou la juge fait ici référence à la CSC elle-même et cite un jugement précédent (comme le montrent les marques typographiques et la référence entre parenthèses), qui explique comment le système juridique anglo-saxon et la manière dont le délit de diffamation est défini dans ce système. Les exemples (108) à (113) illustrent l'idée que ces termes très fréquents et souvent répétés sont imputables à la nature profondément auto-référentielle et intertextuelle du discours juridique, qu'il s'agisse du registre législatif ou du registre judiciaire.

Certains termes particulièrement saillants dans les résultats sont quant à eux liés au thème abordé par un texte en particulier. La fréquence élevée des termes *child*

et *adoption* est par exemple due à la présence de plusieurs textes législatifs dans le corpus EN-Lois qui portent sur le droit familial ou la protection de l'enfance. Le mot *child*, qui, nous l'avons vu au Chapitre 1, a un sens plus large dans ses usages généraux que dans ses usages spécialisés, est surreprésenté dans les deux sous-corpus PLAIN, en particulier EN-CAB. Si ce terme est plus fréquent dans les textes de vulgarisation, c'est sans doute parce que les rédacteur.ices réorientent le discours vers les participants humains en décrivant et expliquant les différentes procédures judiciaires et administratives dont il est question.

- (114) When a **child** is born through surrogacy, the intended parents (IPs) should apply to the family court for a parental order. (...) Without a parental order IPs may not be the **child's** legal parent in the UK unless parenthood is obtained through adoption. (EN-Brochures)
- (115) If you want specific care or treatment as a result of the problem, you may want to approach your GP first, because the Parliamentary and Health Service Ombudsman doesn't normally get involved in someone's ongoing care and treatment.
- Example
- Child** B was an eight-year-old **child** who was discharged from hospital. Staff had missed signs that showed she was seriously ill and sadly she died at home the next day. (EN-CAB)

Cette réorientation du discours d'une expressions abstraite à des informations très concrètes est également visible dans les termes plus spécifiques à chaque sous-corpus individuellement, par exemple *lawyer* dans EN-Brochures et *service* pour EN-CAB. Les rédacteur.ices des textes orientent les justiciable vers des acteurs experts de la justice, qu'il s'agisse de professionnels du milieu judiciaire ou d'organismes spécialisés dans l'aide juridique, comme on le voit dans les exemples suivants :

- (116) If you are thinking about appealing or asking for judicial review, you should talk to a **lawyer** first. (EN-Brochures)
- (117) The Patient Advice and Liaison **Service** (PALS) offers confidential advice, support and information on health-related matters. (EN-CAB)

De même, si les termes qui arrivent en tête dans EN-CA-Résumés et EN-CA-Jugements sont les mêmes (*act* et *court*), le premier contient environ presque deux fois plus d'occurrences de *judge* (pmm) que le second (près de 5000 pmm contre 2478 pmm). Ces différences de fréquence entre EN-CA-Jugements et EN-CA-Résumés laissent supposer que les résumés réorientent également le discours vers les acteurs humains du droit, et en particulier vers la figure centrale du ou de la juge.

- (118) In Mr. Vallentgoed's case, the trial judge erred by not ordering disclosure of the maintenance records. Rowbotham J.A. would have dismissed the Crown appeal in Mr. Vallentgoed's case. With respect to Mr. Gubbins, Rowbotham J.A. would have allowed the appeal in part, as Mr. Gubbins had (later) received disclosure of the complete service records that he had sought. (EN-CA-Jugements)
- (119) The trial **judge** ruled for Mr. Gubbins and ordered the trial stopped until the Crown gave Mr. Gubbins the records. The summary conviction appeal **judge** agreed. (EN-CA-Résumés)
- (120) The Provincial Court **judge** agreed with the Crown. (EN-CA-Résumés) (EN-CA-Résumés)

Ces exemples montrent que les résumés font fréquemment référence aux décisions prises par les différents de juges des différents tribunaux et cours, rappelant le cheminement des affaires à l'intérieur du système judiciaire. L'extrait de jugement (118) suggère que les juges de la CSC mentionnent bien sûr les juges, mais expliquent plus en détail les raisonnements et arguments des cours. De plus, l'expression "Rowbotham J.A." montre qu'ils utilisent le nom propre et l'abréviation J.A. (*justice of appeal*) pour désigner nommément les juges des cours d'appel (ici une juge de la Cour d'Appel d'Alberta), ce qui explique la fréquence plus basse du terme *judge* dans le sous-corpus spécialisé. On note par ailleurs que les textes PLAIN répètent les termes utilisés, comme les textes juridiques : les deux corpus semblent tous deux être marqués par une forte redondance. La densité moyenne des termes étudiés dans chacun des sous-corpus s'expliquerait donc non pas seulement par la présence d'un grand nombre de termes différents, mais par la répétition des mêmes termes ayant un faible degré de spécialisation. Nous avons en effet ici mesuré la densité terminologique en nous fondant sur le nombre de *token* (c'est-à-dire que chaque terme est compté une fois à chaque fois qu'il apparaît dans les corpus, permettant d'inclure les répétitions) et non les *types* (dans ce cas, chaque terme ne serait compté qu'une fois).

Qu'en est-il des termes que l'on peut qualifier comme présentant un haut degré de spécialisation ? Il s'agit des termes n'appartenant pas aux termes-clés ayant une double appartenance (générale et juridique), mais qui sont seulement définis en droit. On constate qu'ils ont une faible densité dans les différents sous-corpus d'EN-LEX. Certains ne concernent que des concepts liés au droit national, comme *certiorari*, spécifique au Canada (dont nous reparlerons plus loin). D'autres termes ne sont pas spécifiquement liés au droit national mais n'apparaissent que dans un seul texte ou une poignée de de textes du corpus juridiques comme par exemple *subpoena*, ou encore *fiduciary*.

- (121) A witness **served with an Australian subpoena** must comply with the **subpoena** (EN-Lois).
- (122) An ad hoc **fiduciary duty** arises where there is an undertaking by the alleged **fiduciary** to act in the best interests of the alleged beneficiaries ; a defined class of beneficiaries vulnerable to the **fiduciary's** control ; and a legal or substantial practical interest of the beneficiaries that stands to be adversely affected by the alleged **fiduciary's** exercise of discretion or control (...) (EN-CA-Jugements)

L'exemple de *subpoena* est issu du texte législatif de Nouvelle-Zélande portant sur la preuve, intitulé *Evidence Act* ; c'est le seul textes du corpus EN-Lois qui contient ce terme. Cet exemple souligne par ailleurs l'importance de l'environnement phraséologique des termes. Ici, pour que les lecteur.ices comprennent la phrase, il est presque aussi important qu'ils comprennent les deux verbes employés comme collocs du terme subpoena, *comply* et *serve*, plutôt qu'ils comprennent le terme lui même. Concernant *fiduciary*, dans l'extrait de jugement, le terme apparaît 310 fois dans le corpus de jugements, mais 299 de ces occurrences proviennent du même texte (bien qu'il apparaisse dans 5 textes en tout), ce qui suggère à nouveau que la distribution des termes est également liée au thème spécifique des textes juridique. Ce terme est utilisé, dans l'exemple, à la fois comme terme simple mais aussi dans des termes complexes où il sert à restreindre le sens d'un nom, par exemple *fiduciary duty*. On voit donc ici que certains termes complexes (poly-lexicaux) sont en fait capturés par notre liste ; une étude plus détaillée des termes complexes permettrait de nuancer nos résultats. Ces termes très spécialisés, qui renvoient à des procédures, documents ou concepts spécifiques, ont par ailleurs une faible densité dans les corpus spécialisés eux-mêmes, et n'apparaissent pas dans les corpus destinés au public. Dans les cas où ils apparaissent, ils font l'objet de stratégies d'explicitation, vers lesquelles nous nous tournerons après avoir donné les résultats pour le français.

7.3 Résultats pour le français

Dans cette section sont présentés les résultats pour le français. Pour le vocabulaire non-approuvé, seuls une dizaine d'exemples trouvés dans différents guides rédactionnels ont été utilisés. Concernant les termes, nous avons choisi d'explorer le passage de FR-CA-Jugements à FR-CA-Résumés grâce à la version française du glossaire de termes juridiques canadiens.

7.3.1 Mots non-approuvés : quelques exemples

Le Tableau 7.6 présente les résultats quantitatifs concernant les éléments lexicaux à éviter selon les guides rédactionnels. On y trouve quelques items lexicaux (résultats donnés pour les lemmes) mais également des prépositions complexes.

TABLEAU 7.6 – Fréquence des mots et expressions "non-approuvées" dans FR-LEX et FR-PLAIN (pmm)

Mot ou expression	FR-Lois	FR-CA-Jugements	FR-DILA	FR-Brochures	FR-CA-Résumés
préalablement	98,4	5,41	46,3	41,2	0,0
conformément	437,3	176,4	22,4	112,3	109,3
solliciter	119,8	95,2	62,4	243,2	40,9
stipuler	17,5	3,3	2,01	0,0	27,3
effectuer	641,9	32,4	541,3	497,7	136,9
procéder	422,9	202,4	154,6	157,2	13,7
requête	158,0	261,9	189,2	89,9	123,0
au sein de	179,6	107,1	56,3	142,2	95,8
dans le cadre de	382,6	149,4	257,6	276,9	150,6

Les données suggèrent des résultats contrastés. Surtout, les éléments lexicaux et phraséologiques semblent être davantage caractéristiques du discours législatif : à part pour *au sein de*, leur place est bien plus haute dans FR-Lois que FR-CA-Jugements. Les sous-corpus correspondant à des brochures ou textes administratifs (FR-DILA et FR-Brochures) semblent également utiliser largement certains de ces éléments lexicaux, comme le verbe *effectuer*, contrairement aux résumés de FR-CA-Jugements, où il a une fréquence quatre fois inférieure aux autres sous-corpus. Il est difficile d'interpréter ces résultats, mais nous pouvons faire l'hypothèse que ces éléments lexicaux, semblent davantage correspondre au style législatif et administratif qu'au style judiciaire. Ces mots ou expressions à éviter ont été extraits de plusieurs sources, dont un guide de rédaction administrative (Collette et al. 2002), ce qui pourrait expliquer ce résultat. Nous avons choisi ici d'illustrer quelques emplois de *procéder*, dans un texte non-simplifié dans un texte simplifié :

- (123) Le président du tribunal **procède** à l'audition, en audience publique, de la personne morale mise en cause et de la victime assistées, le cas échéant, de leur avocat. (FR-Lois)
 Le président du tribunal : entend les parties et **procède** à l'audition des témoins (...). (FR-Brochures)

Dans l'extrait de FR-Brochures, on voit que les rédacteur.ices, qui expliquent le déroulement d'une audience devant le tribunal de police, adoptent la phraséologie du discours spécialisé de départ, puisqu'on voit que le verbe *procéder* est suivi du même prédicat complexe introduit pas une préposition "à + nom prédicatif" (ici *audition*, mais on trouve dans le corpus des occurrences d'autres GN comme *un entretien*, *votre licenciement* ou encore *un tirage au sort*). On trouve une seule occurrence du verbe *procéder* dans FR-CA-Résumés (alors qu'il a une fréquence de 202,4 pmm dans FR-CA-Jugement), imitant là encore la formulation du texte de départ, comme on le voit en comparant cet extrait au texte spécialisés correspondant :

- (124) La cour estime que la Ville a **procédé à une expropriation déguisée** du fait qu'elle a restreint l'usage qui pouvait être fait du lot en modifiant la réglementation de zonage et en installant sur celui-ci des infrastructures d'utilité publique. M. Pichette a accusé la Ville d'avoir **procédé à une "expropriation déguisée"**. (FR-CA-Résumés)

La différence entre les deux textes provient de l'utilisation, dans l'extrait du résumé, de signes typographiques, les guillemets,, pour signaler le terme juridique *expropriation déguisée* (le terme est d'ailleurs suivi par une définition. Concernant les expressions polylexicales à éviter, les exemples ci-dessous illustrent que les sous-corpus de FR-PLAIN utilisent à la même fréquence la préposition complexe *dans le cadre de*.

- (125) **Dans le cadre de** l'habeas corpus, le juge peut se pencher sur la situation dans son ensemble et non pas seulement sur des éléments déjà évalués par le décideur. (FR-CA-Résumés)
- (126) Le journaliste qui a rédigé l'article, M. Marouani, n'a pas soumis d'affidavit **dans le cadre de** la requête. (FR-CA-Jugements)
- (127) **Dans le cadre du** procès pénal, les magistrats du parquet (le procureur général, le procureur de la République ou les substituts du procureur), et non la victime, demandent la condamnation de l'auteur de l'infraction au nom de la société. (FR-Brochures)
- (128) Le témoin prêtant serment **dans le cadre d'**une instruction, un faux témoignage est considéré comme un délit. (FR-Brochures)

Cette préposition complexe est utilisée aussi bien dans les textes spécialisés que dans ceux s'adressant au public pour préciser le contexte judiciaire (instruction, procès pénal) ou la procédure dont il est question (requête, habeas corpus). Comme pour l'anglais, on peut en conclure que les textes de diffusion du droit utilisent au moins partiellement certains éléments lexicaux et grammaticaux caractéristiques

du discours juridique : on trouve une trace du discours original dans ces textes. A nouveau on peut faire l'hypothèse que pour les rédacteur.ices de ces textes, il est possible que le lectorat soit familier de ces mots ou expressions, et que c'est pour cela qu'elles sont employées. La faible taille de l'échantillon ne permet pas de tirer des conclusions plus générales sur la densité d'éléments lexicaux considérés complexes dans FR-LEX et FR-PLAIN. La création d'une liste ad hoc serait nécessaire pour explorer davantage cette question.

7.3.2 Densité des termes dans FR-CA-Jugements et FR-CA-Résumés

La Tableau 7.7 présente la fréquence des termes (en %) du glossaire dans les motifs de jugement francophones de la CS du Canada et leurs résumés.

TABLEAU 7.7 – Proportion des termes dans le discours judiciaire et sa simplification (FR-CA) par rapport au nombre total de mots

Corpus	Lectorat visé	Proportion des termes (%)
FR-CA-Jugements	Spécialisé	4,9
FR-CA-Résumés	Non-expert	3,6

Ce tableau suggère que la proportion de termes est légèrement plus faible dans FR-CA-Résumés que dans FR-CA-Jugements, bien que cette différence reste faible par rapport aux différences entre certains mots lexicaux considérés complexes dans les deux sous-corpus. Dans le Tableau 7.8 sont présentés les termes les plus fréquents d'après AntWord Profiler, ainsi que leur fréquence pour un million de mots.

Comme pour le corpus parallèle anglophone, les termes les plus fréquents sont les termes centraux de la procédure judiciaire, qui présentent un faible degré de spécialisation : *loi*, *droit*, *preuve*, *juge* et *cour*. Ce dernier est fréquent en particulier, car la Cour fait référence à elle-même ainsi qu'aux cours inférieures. Deux termes qui diffèrent dans chacun des deux corpus sont cependant intéressants. FR-CA-Résumés contient 1059,8 occurrences pmm de *peine* (qui a une fréquence à peu près similaire dans FR-CA-Jugements) et 976,1 occurrence pmm de *coupable* (contre seulement 314,5 dans FR-CA-Jugements), tandis que les termes *décision* et *charte* sont parmi les termes les plus saillants dans FR-CA-Jugements. Nous présentons d'abord des exemples des deux premiers :

TABLEAU 7.8 – Termes les plus fréquents dans le discours judiciaire francophone avant et après simplification

(a)		(b)	
Terme	Fréquence dans FR-CA - Jugements (pmm)	Terme	Fréquence dans FR-CA-Résumés (pmm)
loi	3 786,4	cour	5 884,4
juge	3 683,4	juge	4 810,7
droit	3 587,6	droit	3 165,3
cour	3 295,6	loi	2 691,2
appel	2 152,0	appel	1 101,6
décision	1 944,7	peine	1059,8
charte	1 513,1	preuve	1017,9
preuve	1 383,4	coupable	976,1

- (129) La décision a fait ressortir qu'une **peine** devait être représentative à la fois de la gravité du crime et du degré de responsabilité de la personne.
- (130) Le juge du procès a déclaré R.V. coupable et l'a condamné à une **peine** d'emprisonnement de quatre ans.
- (131) La Cour suprême a jugé que la personne reconnue **coupable** a droit soit à la **peine** qui s'appliquait quand elle a commis son crime, soit à celle qui s'appliquait quand elle a été condamnée. (FR-CA-Résumés)

On peut à nouveau supposer que dans le corpus destiné au public, l'accent est mis sur les référents humains et les situations juridiques (condamnations, par exemple) qui sont susceptibles de concerner directement et d'intéresser les justiciables, ce qui se traduit par une plus grande fréquence des termes qui y font référence. Dans FR-CA-Jugements les deux termes saillants renvoient à des textes ou concepts plus abstraits, en particulier *charte*, qui désigne la Charte canadienne des droits et libertés, texte fondamental des droits civiques et humains au Canada, et *décision*, qui fait référence aux jugements antérieurs :

- (132) Si le tribunal est convaincu qu'il y a eu atteinte à un droit, la seconde étape de l'analyse consiste à décider si l'État peut justifier cette atteinte au regard de l'article premier de la **Charte**.
- (133) La juge du procès a conclu que des **décisions** antérieures avaient déjà établi l'existence d'une telle obligation.

C'est le fondement même du rôle de la CSC qui explique la fréquence et l'utilisation de ces termes, par l'expression des questions de droit portant sur l'interprétation

et l'application de la Charte, ainsi que des références à la jurisprudence et aux décisions des cours inférieures (et de la Cour elle-même dans des affaires passées). Il semble cohérent que les motifs de jugement fassent référence à ces autres textes, tandis que les résumés expliquent les faits et éléments judiciaires de l'affaire qui ont des conséquences pour l'ensemble des justiciables.

Si l'on se penche à présent sur les termes ayant un plus haut degré de spécialisation, on remarque qu'ils sont peu fréquents dans le corpus de jugements - par exemple 62 occurrences pmm pour *affidavit*, et ont une fréquence faible voire nulle dans le corpus de résumés (c'est le cas pour *affidavit*, qui n'apparaît pas dans FR-CA-Résumés). Comme pour l'anglais, on en déduit que, au moins pour les termes mono-lexicaux, ce sont les termes ayant un faible degré de spécialisation qui ont une plus grande densité dans les deux sous-corpus. Concernant les termes poly-lexicaux qui n'ont pas été capturés par l'outil, une recherche en corpus révèle par exemple la présence dans FR-CA-Jugements de plusieurs termes en latin sous formes de locutions, comme *lex loci delicti* (43,3 pmm) et *forum non conveniens* (129,9 pmm).

- (134) Lorsque la compétence se fonde sur le lieu du délit, le droit applicable suivant la **lex loci delicti** pointera effectivement vers le ressort en question. Cela ne veut pas dire qu'il convient d'accorder peu de poids au facteur du droit applicable dans l'analyse relative au **forum non conveniens**. (FR-CA-Jugements)

Ces termes, qui désignent des méthodes et concepts de l'analyse juridique, soit sont absents dans les résumés en style clair, soit font l'objet d'une stratégie d'explicitation. Nous nous tournons maintenant vers ces stratégies.

7.4 Analyse qualitative : Stratégies d'explicitation des mots non-approuvés et des termes spécialisés dans le corpus PLAIN

Après avoir analysé les différences quantitatives des mots et termes spécifiques à chacun des discours, nous présentons brièvement quelques stratégies explicatives des termes ou items lexicaux complexes présentes dans le corpus EN-PLAIN, afin d'allier une étude qualitative aux résultats quantitatifs.

7.4.1 Degrés de spécialisation des termes, termes-clés et dé-terminologisation

Comme nous venons de le voir, les résultats quantitatifs suggèrent que les termes les plus fréquents sont des termes présentant un bas degré de spécialisation dans le corpus LEX et dans le corpus PLAIN. Nous présentons ci-dessous des exemples de ces termes en contexte, tirés du corpus EN-PLAIN.

TABLEAU 7.9 – Exemple de termes "peu spécialisés" dans EN-PLAIN

Terme	Exemple dans le corpus EN-PLAIN
court	After you have given evidence you will be excused from the court . (EN-Brochures)
appeal	He or she will either allow your appeal or dismiss (refuse) your appeal . (EN-Brochures)
act	The Matsons and the Andrewses were arguing the Indian Act itself was discriminatory. (EN-CA-Résumés)
law	The defence has to be based on facts and law . (EN-CA-Résumés)

Dans ces exemples anglophones, ces termes ne font pas l'objet d'une définition ou d'une reformulation. Dans la deuxième ligne du tableau, ce n'est pas le terme *appeal* qui est explicité, mais le verbe qui entre en collocation avec lui, *dismiss*, qui est reformulé entre parenthèses à l'aide du verbe *refuse*. Pour les rédacteur.ices, il semble donc que ce ne n'est pas le terme qui a besoin d'être explicité, mais la phraséologie et les collocations spécifiques qui forment leur environnement.

Qu'en est-il pour le français? Nous avons choisi les exemples comportant des termes les plus fréquents dans les résumés des motifs de jugements francophones de la CSC mais aussi dans les autres sous-corpus francophones, car ils sont également des termes-clés fondamentaux dans le système juridique francophone.

Nous pouvons en déduire que ces termes, qui sont au cœur du domaine du droit, à la fois pour ces acteurs et actrices et pour ses usagers, n'ont pas besoin d'être explicité car ils ont subi un processus de déterminologisation, c'est-à-dire qu'ils sont "passés" dans le domaine général" (Condamines et Picton 2014). Leur ancrage social est très fort, puisqu'ils sont utilisés par de nombreux usagers et professionnel.les, et leur origine historique est très ancienne, ce qui a facilité leur intégration à des usages du domaine général; Cornu évoque les emplois métaphoriques de ces termes-clés (2005 : 72), notamment dans des expressions idiomatiques lexicalisées. On peut par exemple penser à *as sober as a judge* en anglais, ou encore le *tribunal populaire des*

TABLEAU 7.10 – Exemple de termes "peu spécialisés" dans FR-PLAIN

Terme	Exemple dans le corpus FR-PLAIN
cour	Le greffier [sic] assiste la cour , note le déroulement des débats, met en forme et authentifie la décision. (FR-Brochures)
juge	Le juge peut rendre sa décision le jour même de l'audience ou bien mettre la décision en délibéré, c'est-à-dire prévoir une autre date pour faire connaître son jugement. (FR-DILA-Brochures)
droit	En laissant le jury entendre la preuve, la juge a commis une erreur de droit . (FR-CA-Résumés)
loi	Le gouvernement fédéral a le pouvoir de faire des lois sur les crimes. (FR-CA-Résumés) Pour savoir si vous avez droit à l'aide juridictionnelle et calculer vos droits , vous pouvez utiliser le simulateur suivant. (FR-DILA)

réseaux sociaux dans la presse française. Ces emplois figurés attestent du glissement et de l'intégration de ces termes dans les usages généraux de la langue. Ils expliquent probablement pourquoi ces termes n'ont donc pas besoin d'être explicités dans le corpus PLAIN, car ils seront probablement reconnus par le lectorat.

Cependant, la plupart des termes du corpus sont des termes d'appartenance juridique simple, c'est-à-dire n'ayant de définition qu'en droit, ou des termes ayant un sens très différent de leur sens général dans des textes spécialisés ; ils nécessitent donc une définition ou reformulation pour être compréhensibles par les non-expert.es. Nous nous tournons maintenant les stratégies d'explication de ces termes.

7.4.2 Définitions, reformulations et utilisation des alternatives claires et simples

Nous nous intéressons ici aux termes du glossaire étudiés dans la section précédente, mais ouvrons aussi l'analyse à d'autres termes juridiques présents dans le corpus EN-PLAIN. Cela permet d'inclure de nombreux termes qui n'ont pas été capturés par la liste utilisée, notamment les termes complexes (polylexicaux).

La première stratégie employée est la définition après une première mention du terme ou du mot, dans une phrase contenant la copule (*be* ou *être*) et une proposition (soulignée dans les exemples) qui définit le terme ou le mot jugé complexe.

(135) An **adjournment** is // when the judge puts the case on hold. (EN-CAB)

- (136) Mr. Wall asked the Alberta Court of Queen's Bench to **review** the decision. **Judicial review is // when courts look at tribunals decisions //** to make sure they are proper. (EN-CA-Résumés)
- (137) L'**autorité parentale** est un ensemble de droits et de devoirs ayant pour finalité l'intérêt de l'enfant mineur, // afin de le protéger dans sa sécurité, sa santé et sa moralité, // pour assurer son éducation et permettre son développement.
- (138) La **prestation de compensation du handicap** est une aide financière versée par le Conseil départemental, destinée // à financer les besoins liés à la perte d'autonomie des personnes handicapées. (FR-Brochures)

Dans l'exemple (135), la définition de *adjournment* apparaît dans une proposition subordonnée en *when*, tandis que *judicial review* dans l'exemple (136) est défini par trois propositions imbriquées, dont une proposition en *to* qui explicite la fonction du contrôle judiciaire. Dans les exemples en français, la proposition contenant *être* est suivie d'une proposition en *afin* ou d'une proposition réduite (*small clause*) au passif (exemple (138) *destinée à*), qui servent à également expliquer la fonction du concept juridique ou de la procédure dont il est question, ici les notions de contrôle judiciaire et d'autorité parentale. Une autre stratégie d'explicitation des termes spécialisé consiste à paraphraser le terme en question. Dans le corpus EN-PLAIN, l'étude en contexte révèle que certains termes ou items lexicaux complexes sont paraphrasés par l'alternative en PL qui est indiquée par PLC, en rouge dans les exemples suivants :

- (139) **Statutory** employment rights are given to workers and employees by Acts of Parliament rather than through a contract of employment. Asking for what you're entitled to **by law is called** asserting a statutory right. (EN-CAB)
- (140) If these points could reasonably have affected the original judge's decision, another judge has to review the whole decision from the beginning (not just **looking at** whether it had a reasonable basis). (EN-CA-Résumés)
- (141) Mr. Calnen was charged with second-degree murder (**intentionally causing Ms. Jordan's death**). He was also charged with indecent interference with human remains (**damaging or disrespecting a dead person's body**).
- (142) Part of deciding the appropriate place for a case to be heard is deciding which laws should apply ("**choice of law**").

La reformulation se fait dans certains cas sans commentaire méta-linguistique explicite, mais seulement à l'aide de signes typographiques, comme on le voit dans les exemples ci-dessus où *statutory* est explicité par *by law* (139) et le verbe *review* est reformulé entre parenthèses à l'aide de son équivalent clair et simple *look at again*

(140) suggéré dans la liste "*A Z of alternative words*". Parfois, à l'inverse, c'est le terme qui est donné après la définition entre parenthèses et guillemets, comme pour *choice of law*. Mais la paraphrase ou définition du terme est aussi souvent introduite par des séquences métatextuelles figées ou semi-figées comme *that is*, introduisant une ou plusieurs propositions qui servent à définir le terme :

- (143) The Superior Court judge ruled for the Crown and granted certiorari (**that is**, canceled the provincial court judge's order). (EN-CA-Résumés)
- (144) You can claim compensation only if (...) your treatment was carried out negligently, **that is** , the care you received fell below medically acceptable standards, and this directly caused your injury. (EN-CAB)

L'une des séquences figées pour introduire un terme en anglais est *this is called*, d'une fréquence de 277,0 pmm (dont on remarquera qu'elle est au passif, ce dont nous reparlerons dans le chapitre suivant).

- (145) A State may also lodge an application against another State Party to the Convention ; **this is called** an inter-State application. (EN-Brochures)
- (146) If you don't have a religious belief, there's another card where you still swear that everything you say will be the truth (**this is called** an affirmation). (EN-CAB)

En français, on trouve l'équivalent dans EN-CA-Résumés avec la séquence *C'est ce qu'on appelle* + [terme entre guillemets], comme on le voit ci-dessous :

- (147) Lorsqu'ils rendent des décisions, les tribunaux administratifs doivent traiter tout le monde équitablement et décider de façon impartiale. **C'est ce qu'on appelle** l' "équité procédurale".
- (148) Lorsque de nombreuses personnes ont le même problème juridique, elles peuvent décider d'unir leurs forces et de poursuivre en justice en tant que groupe. C'est ce qu'on appelle un "recours collectif". (L'adjectif "collectif" renvoie au groupe, il s'agit donc d'un recours intenté en groupe.)

Dans le corpus FR-DILA, une autre séquences fréquentes pour reformuler un terme et l'expliquer est *c'est-à-dire* (équivalent de *that is* vu plus haut dans EN-PLAIN), locution suivie d'une proposition ou d'un GN, qui apparaît 199,2 fois par million de mots.

- (149) Le juge peut aussi prendre une décision d'incompétence, **c'est-à-dire** qu'il estime que le litige ne relève pas de ce tribunal de grande instance.
- (150) Si le projet d'adoption concerne un pupille de l'État, **c'est-à-dire** un enfant remis à un organisme autorisé pour l'adoption ou un enfant étranger, le couple doit obtenir au préalable un agrément.

De plus, dans ce sous-corpus, les rédacteur.ices utilisent des stratégies systématiques pour définir certains termes. Ainsi trouve-t-on 160 occurrences (322,0 pmm) de la question "De quoi s'agit-il?", qui sert de véritable élément structurant en introduisant les termes dont il est question dans le texte.

(151) **De quoi s'agit-il ?**

Cas général

La perquisition est la fouille d'un lieu en vue d'y trouver des preuves d'une infraction.

Les questions/réponses renforcent le dialogisme entre les rédacteur.ices et le lectorat non-spécialisé (Preite 2018). Surtout, cette stratégie, ajoutée aux reformulations des termes, montre que les rédacteur.ices anticipent les difficultés et problèmes de compréhension de leurs interlocuteur.ices :

(...) in written discourse in which the text is a finished product, reformulations can be seen as an anticipation of an an attempts to pre-empt obstacles to understanding. (Turnbull 2018)

Les textes du corpus PLAIN utilisent donc des méthodes de reformulation en accord avec la littérature sur la vulgarisation et la médiation. L'étude des contextes révèle cependant des différences entre les sous-corpus, par exemple entre FR-CA-Résumés(canadien) et FR-DILA (français). Les exemples en français canadien contiennent davantage de formes verbales introduisant une situation par *lorsque*, suivi d'une explication des actions des tribunaux ou des parties grâce à des formes verbales. Dans FR-DILA, on trouve également ce type d'explicitation (notamment dans la définition de la décision d'incompétence), mais de nombreuses définitions contiennent aussi des groupes nominaux post-modifiés. En particulier, la le terme *perquisition* est défini par une nominalisation (*la fouille*) et non d'une proposition (par exemple *lorsque la police fouille un lieu, etc*). De même, le terme *pupille de l'Etat* est défini via une post-modification du nom tête *enfant*, suivie par une cascade de noms modifiés par des propositions réduites et des groupes prépositionnels (*pour l'adoption*). Cette différence pourrait être due à l'application des principes du style clair dans les résumés de la CS en français, qui privilégieraient des formes nominales.

Le corpus FR-DILA, issu d'une administration française où ces principes ne font pas partie de la politique linguistique, semble rester plus proche du style législatif sur lequel se fondent les textes, et donc privilégier l'emploi de formes nominales. La voix adoptée par les rédacteur.ices du corpus FR-DILA resterait donc plus proche de la voix des législateur.ices. Cette interprétation semble se confirmer lorsqu'on compare la définition de certains termes dans FR-DILA au texte de loi d'origine. Confrontons la définition du harcèlement sexuel dans FR-DILA et FR-Lois :

- (152) (...) des faits (...) de harcèlement sexuel, constitué par **des propos ou comportements à connotation sexuelle** répétés qui soit **portent atteinte à sa dignité en raison de leur caractère dégradant ou humiliant, soit créent à son encontre une situation intimidante, hostile ou offensante.** (FR-Lois)
- (153) De quoi s'agit-il? Le harcèlement sexuel se caractérise par le fait d'imposer à une personne, de façon répétée, **des propos ou comportements à connotation sexuelle ou sexiste, qui : portent atteinte à sa dignité en raison de leur caractère dégradant ou humiliant, ou créent à son encontre une situation intimidante, hostile ou offensante.** (FR-DILA)

En orange sont indiqués les éléments de la définition qui sont exactement les mêmes dans le texte législatif et dans le texte destiné aux justiciables. Si la question *De quoi s'agit-il?* annonce une explicitation dans (153), on voit que les rédacteur.ices modifient à peine le texte spécialisé d'origine présenté en orange dans (152). Le mot *personne* est utilisé pour désigner la victime, plutôt que le nom plus courant *quelqu'un*, de même que *à son encontre* est conservé et non remplacé par *pour elle*. Les adjectifs *intimidante, hostile ou offensante* ne sont pas explicités. Cet exemple n'est pas un phénomène isolé : il a été porté à notre connaissance en conférence et lors de discussion avec les usagers que les articles de loi étaient peu voire pas du tout modifiés avant d'être publiés sur le site *Service Public*. Cela expliquerait pourquoi la voix discursive semble si proche de celle du discours spécialisé par rapport à FR-CA-Résumés.

Les stratégies définitoires sont parfois accompagnées d'exemples plus ou moins élaborés pour éclairer un concept juridique, qui permettent de mettre au jour des différences cette fois entre des institutions francophones et anglophones. Nous prenons un exemple de FR-DILA et EN-CAB.

- (154) La **diffamation non publique** concerne les allégations prononcées (...) devant un cercle restreint de personnes partageant les mêmes intérêts, que la victime soit présente ou non. Les personnes témoins ont toutes un même lien entre elles. Ce lien peut être professionnel,

personnel... Par exemple, une injure lancée lors d'un comité d'entreprise est non publique, car prononcée devant un nombre restreint de personnes appartenant à une même instance. (FR-DILA)

- (155) It's unlawful if someone discriminates against you because they think you're of a particular sexual orientation even though you're not. This is called **discrimination by perception**.
 Example : You're heterosexual and about to start your degree at University. You apply for a place in an all-female hall of residence. The accommodation officer thinks you're gay and refuses to offer you a place as she thinks the other students are going to feel uncomfortable living with a lesbian woman. This is unlawful discrimination because of sexual orientation by perception. (EN-CAB)

Dans l'extrait (154), écrit par une administration française, la définition de *difamation publique* est donnée puis illustrée par un exemple laconique, baptisé tout simplement *exemplification* en analyse de discours (Turnbull 2018). L'exemple donné contient un élément lexical qui pourrait être considéré comme complexe, (*restreint*), et en termes syntaxiques, contient un passif non périphrastique (*prononcée*) modifié par plusieurs GN, ce qui résulte en un exemple relativement peu précis et surtout impersonnel.

Dans l'extrait (155) tiré de EN-CAB, à l'inverse, le long exemple correspond à un scénario, selon le terme employé Turnbull (2018), c'est-à-dire que qu'il exprime une situation plus complexe afin d'illustrer très précisément un cas auquel s'applique le terme *discrimination by perception*. L'exemple comprend également une forme de complexité syntaxique, cette fois liée à l'imbrication de six propositions hypotactiques (de *The accomodation officer* à *a lesbian woman*). Le ton est indéniablement différent de l'extrait francophone, notamment grâce à l'emploi des pronoms de deuxième personne, des formes contractées et la description d'actions concrètes. L'oralité et l'informalité de l'extrait anglais évoque une dimension interpersonnelle plus forte, tandis que la contextualisation narrative plus longue laisse à penser qu'un plus grand effort de concrétisation (*concretization*, Turnbull (2018)) des connaissances juridiques est fait dans EN-CAB que dans FR-DILA.

Malgré ces différences entre les sous-corpus, ces analyses qualitatives confirment les résultats concernant la complexité syntaxique (*intricacy*) décrites au Chapitre 6 : l'explicitation du lexique et des termes est réalisée par l'ajout de propositions enchâssées, subordonnées ou juxtaposées qui viennent élaborer le sens, ce qui résulte en une plus grande complexité syntaxique.

7.4.3 Quelle influence du genre spécialisé source sur les stratégies explicatives ?

L'étude qualitative des contextes révèle également des spécificités liées au genre discursif de départ. En particulier, comme nous l'avons vu au Chapitre 1, les jugements de la Cour suprême sont rendus oralement par le Juge en chef et éventuellement par l'opposition. Or, dans les sous-corpus de CA-Résumés, on trouve non seulement des définitions des termes mais aussi des éléments qui rappellent cette dimension orale du texte d'origine, comme on le voit ci-dessous :

- (156) *Certiorari* (**pronounced different ways, including ser-tee-oh-RARR-ee**) is a court's power to cancel another court's decision because of an obvious mistake.
- (157) He applied for habeas corpus (pronounced "HAY-bee-us KOR-pus") in 2016. Habeas corpus is an old and important legal concept, dating back many centuries. It means "produce the body" in Latin. Taking someone's freedom away should be a last resort, and habeas corpus guarantees it won't happen illegally. (EN-CA-Résumés)

Les termes latins ci-dessus sont donc définis, traduits et explicités, et on précise également leur prononciation, non pas en alphabet phonétique mais grâce à une sorte de translittération phonologique en alphabet latin, car les rédacteur.ices supposent que le lectorat visé ne connaîtra pas nécessairement l'alphabet phonétique. Cette dimension multimodale est liée au mode discursif du genre spécialisé de départ, les jugements étant avant tout oraux. La prononciation est précisée afin que les justiciables puissent reconnaître le terme s'ils l'entendent lors d'un procès, ou en regardant la diffusion ou rediffusion d'une audience, les audiences étant disponibles sous forme vidéo sur le site de la Cour Suprême du Canada. Le sous-corpus FR-CA-Résumés contient également des indications sur la prononciation pour *certiorari*.

- (158) Le *certiorari* (**qui se prononce « cerciorari »**) désigne le pouvoir d'un tribunal d'annuler la décision d'un autre tribunal en raison d'une erreur flagrante.
- (159) Il a présenté une demande d' habeas corpus en 2016. L'habeas corpus est un important concept juridique, vieux de plusieurs siècles. Il signifie, en latin, "que tu aies le corps pour le soumettre".

On remarque que la prononciation est précisée pour *certiorari* mais pas pour *habeas corpus*. Cette différence entre le français et l'anglais souligne aussi la dimension subjective de certains choix d'explicitation : le/la rédacteur.ice du texte en anglais a

jugé nécessaire d'indiquer la prononciation d'*habeas corpus*, tandis que le traducteur a fait un choix différent, considérant sans doute que la prononciation et l'orthographe étaient plus proches en français que pour *certiorari*. On notera par ailleurs que cette indication de la prononciation n'est pas réservée aux termes juridiques en latin, comme on le voit dans l'exemple ci-dessous, où c'est l'adjectif *indictable* dont la prononciation est précisée :

- (160) The Criminal Code says that anyone found guilty of a crime has to pay a "victim surcharge." The amount depends on the type of crime. (...) If there is no fine, it is \$100 for every summary conviction count and \$200 for every indictable (**pronounced "in-DYE-ta-bull"**) count.

Dans ce sous-corpus, il semble donc qu'il y ait bien une influence du genre discursif, en tout cas du mode discursif de départ, sur les stratégies d'explicitation des termes, et les définitions textuelles ne sont pas les uniques points d'entrée dans la terminologie judiciaire.

Concernant les textes de législatifs, dans les pays anglo-saxons, leur structure générale comporte une section liminaire qui définit les termes utilisés dans le reste de la loi, dont nous présentons un exemple dans la Figure 7.4.

FIGURE 7.4 – Définition des termes dans le discours législatif néo-zélandais (EN-Lois)

4 Interpretation

In this Act, unless the context otherwise requires,—

absolute discretion has the meaning given to it in [section 11](#)

address for service has the meaning given to it by [section 387](#)

administrative error, in relation to the granting of a visa or entry permission, has the meaning given to it in [section 8](#)

airport has the meaning given to it in [section 2](#) of the Airport Authorities Act 1966

appeal on humanitarian grounds means an appeal to the Tribunal against liability for deportation on the grounds set out in [section 207](#)

appeal on the facts means an appeal against liability for deportation on a ground set out in [section 202](#)

appeals body means 1 or more of the following bodies established or continued under the former Act, as the case may be:

- (a) the Residence Review Board;
- (b) the Removal Review Authority;
- (c) the Refugee Status Appeals Authority;
- (d) the Deportation Review Tribunal

approved system means a system, including an electronic system, approved by the chief executive for the purpose of—

- (a) providing information to the chief executive under [section 96](#); or
- (b) giving notice under [section 97\(2\)](#) of a decision made under [section 97\(1\)](#); or

Cette section est appelée *Interpretation*. Elle vise donc explicitement à guider les juges et magistrat.es dans l'interprétation du texte législatif, rappelant la dimension cruciale de la définition des termes et mots spécifiques au texte de loi en question, en particulier permettre aux juges de comprendre l'intention des législateur.ices dans leur analyse des textes législatifs³. Or, il semble qu'on retrouve des traces de cette section dans certains textes du sous-corpus EN-Brochures : au début du texte portant sur cette même procédure d'appel, on trouve un glossaire comprenant une liste de termes, qui présente les définitions utilisées dans la suite du document, notamment celle de *appellant*.

FIGURE 7.5 – Définition des termes au début d'un texte du sous-corpus EN-Brochures

Explanation of terms

appellant	Someone making an appeal to the Immigration and Protection Tribunal against an Immigration New Zealand decision (or in some cases, a decision made by the Minister of Immigration).
certified copy of a document	A true copy, approved by someone legally allowed to approve it. If you are in New Zealand, certified copies must be approved by a lawyer, notary public, Justice of the Peace or court official.

Cette liste en début de texte fait fortement penser à la section *Interpretation* au début du texte spécialisé correspondant. L'influence du style juridique se voit donc également dans la macro-structure textuelle dans certains textes du corpus PLAIN. Bien sûr, on peut également voir là une stratégie typique de la vulgarisation. En français, des sections intitulées *Glossaire* ou *Lexique* sont parfois incluses dans les brochures qui portent sur des points juridiques précis (par exemple les droits des détenu.e.s), mais plutôt à la fin du texte, ce qui correspond à une stratégie plus généralement liée à la diffusion des connaissances spécialisées - ainsi Ferraresi (2019) note-t-il l'utilisation de glossaires pour expliciter les termes scientifiques liés à la sécurité alimentaire.

3. L'interprétation de l'intention et l'importance de la formulation des textes législatif est soulignée dans les textes de la CSC, par ex : *This interpretation is consistent with the careful wording of the provision, which expresses Parliament's intention to disclaim liability for matters falling under provincial responsibility.* (EN-CA-Jugements)

7.5 Résumé du chapitre 7

Dans ce chapitre, nous nous sommes intéressés au lexique et à la terminologie dans le passage du style juridique au style clair. Les résultats concernant les items lexicaux non-recommandés montrent qu'il n'y a pas d'adhésion totale aux recommandations en style clair, puisque les mots à éviter selon le PLM sont bien présents dans le corpus EN-PLAIN. Les résultats quantitatifs suggèrent cependant qu'il y a une forte diminution (deux à trois fois moins) des mots considérés complexes par les institutions promotrices du style clair. Les résultats quantitatifs sont cependant à nuancer du fait de la nature parfois ambiguë, sémantiquement ou morphologiquement, de certains mots de la liste de vocabulaire. Concernant les termes, les résultats sont moins marqués et la densité terminologique ne semble pas différencier nettement les deux registres, parce que les corpus LEX comme PLAIN, en anglais et en français, sont denses en termes-clés qui sont répétés à travers les textes. Ces termes sont nécessaires pour transférer les connaissances spécialisées aux justiciables, et viennent confirmer les conclusions de certaines études qui suggèrent que les termes dans des documents juridiques ne peuvent être simplifiés que jusqu'à un certain point (Rossetti et al. 2020).

Notre étude des stratégies d'explicitation suggère quant à elle que les définitions et reformulations sont typiques du passage d'un discours spécialisé à un discours destiné aux non-experts. Il s'accompagne d'une exacerbation de la dimension métalinguistique (Jacobi 1985), avec des segments répétés ou des signes typographiques qui permettent de reformuler explicitement les termes. Ces analyses qualitatives mettent par ailleurs au jour des différences entre les sous-corpus francophones et anglophones destinés aux non-spécialistes.

Enfin, cette étude quantitative et qualitative permet d'affirmer le caractère relatif de notion de complexité linguistique, cette fois non pas du point de vue des personnes qui reçoivent le texte, mais celles qui le produisent. Les résultats quantitatifs aussi bien que les contextes révèlent une forme de subjectivité dans les choix explicatifs, possiblement liée à l'institution ou aux auteur.ices individuel.les. Si les éditeur.ices de PLC jugent certains mots complexes et préconisent qu'ils soient remplacés, les rédacteur.ices du corpus PLAIN jugent eux essentiel d'employer le lexique et la terminologie de l'institution judiciaire, législative ou administrative dans certains cas.

Chapitre 8

Fréquence et usage du passif dans les corpus LEX et PLAIN

Dans ce chapitre, nous cherchons à vérifier l'hypothèse suivante : les textes du corpus PLAIN contiennent une faible fréquence de constructions à la voix passive, tandis que ceux du corpus LEX, en particulier les textes législatifs, en contiennent une grande proportion. Le passif est en effet considéré comme particulièrement caractéristique du discours législatif, bien qu'il ait évolué dans les dernières décennies (comme vu au Chapitre 1). Concernant le corpus PLAIN, nous souhaitons encore mesurer le degré d'adhérence des textes de diffusion du droit aux recommandations rédactionnelles, puisque, comme nous l'avons vu dans les Chapitres 3 et 5, éviter le passif est l'une des recommandations les plus fréquentes dans les guides rédactionnels pour rédiger en style clair. Nous proposons ensuite une interprétation quant aux fonctions discursives du passif observées dans les deux registres représentés par les corpus LEX et PLAIN, mais également quant aux différences éventuelles entre les genres discursifs.

En particulier, notre objectif est d'observer si le passif apparaît dans des schémas lexico-grammaticaux (schémas LG). Comme nous l'avons vu au Chapitre 1, les schémas LG sont des structures qui présentent un certain degré de figement, où certains éléments continus ou discontinus servent de pivot (c'est-à-dire qu'ils constituent les éléments récurrents, figés) et dont le contexte étendu révèle des régularités au niveau sémantique (par exemple, le référent du sujet est toujours un acteur institutionnel) et/ou grammatical (par exemple, une structure passive ou infinitive). Nous utilisons les conventions typographiques suivantes pour décrire les schémas LG observés :

— < début du schéma LG

- > fin du schéma LG
- *italique* : pivot (élément obligatoire)

crochets : élément (grammatical) variable mais obligatoire

8.1 Résultats quantitatifs : fréquence du passif canonique dans les corpus LEX et PLAIN

8.1.1 Fréquence du passif canonique dans les corpus anglophones

La fréquence des constructions passives prototypiques (lemme "être" + Participe passé) dans les corpus LEX et PLAIN a été calculée par sous-corpus, afin de faire apparaître d'éventuelles différences génériques. Elle est donc présentée par sous-corpus, dans le Tableau 8.2 pour les textes destinés au grand public du corpus PLAIN et dans le Tableau 8.1 pour les textes spécialisés du corpus LEX.

Cependant le test de significativité statistique du χ^2 a été calculé pour l'ensemble des occurrences de passif par rapport au nombre de groupes verbaux conjugués pour chaque sous-corpus. Les valeurs de χ^2 obtenues sur les différents corpus montrent que les textes du corpus EN-LEX contiennent significativement plus de constructions passives que ceux du corpus EN-PLAIN (¹).

TABLEAU 8.1 – Fréquence de la voix passive dans le corpus EN-LEX

Sous-corpus	Nombre brut de constructions passives	% de constructions passives par rapport au nombre total de mots (tokens)	Nombre de groupes verbaux conjugués	% de constructions passives par rapport au nombre de groupes verbaux conjugués
EN-Lois	29951	1,3	109822	27,3
EN-CA-Jugements	7264	1	48314	15,0

Le sous-corpus de textes législatifs EN-Lois et le corpus de jugements de la CSC contiennent à peu près le même nombre de constructions passives par rapport au nombre total de mots dans le corpus. Cependant, rapporté au nombre de groupes verbaux, on voit que le corpus de lois contient plus de 10% de passifs canoniques que le corpus de jugements. Cette fréquence relative des passifs canoniques dans EN-Lois n'est pas surprenante, puisqu'elle est cohérente avec les conclusions de travaux récents sur le discours législatif anglophone. Notre résultat correspond en particulier à

1. $\chi^2(1, N = 434969) = 2984167, p\text{-value} < .00001$

la proportion décrite par Williams dans son étude de l'évolution des textes législatifs en anglais, dans laquelle il avait rapporté que 26% des verbes transitifs de son corpus de lois de 2010 étaient à la voix passive Williams (2013). La différence entre les deux sous-corpus EN-Lois et EN-CA-Jugements est plus surprenante, et pourrait refléter une différence entre les genres discursifs législatif et judiciaire. Cette différence est cependant à considérer avec précaution, puisque ces résultats ne prennent pas en compte les passifs non-canoniques, par exemple sans auxiliaire.

TABLEAU 8.2 – Fréquence de la voix passive dans le corpus EN-PLAIN

Sous-corpus	Nombre brut de constructions passives	% de constructions passives par million de mots (<i>tokens</i>)	Nombre de groupes verbaux conjugués	% de constructions passives par rapport au nombre de groupes verbaux conjugués
EN-CAB	9574	1,2	70580	13,6
EN-Brochures	3013	1,3	17486	17,2
EN-CA-Résumés	711	1,3	5680	12,5

Les résultats quantitatifs décrits ci-dessus suggèrent une nette diminution du nombre de constructions passives dans certains sous-corpus PLAIN par rapport à leur version du corpus LEX, non simplifiée. Cette différence est par exemple très nette entre le corpus EN-Lois et le corpus EN-CAB, puisqu'on passe de 23,6% à 13,7% de constructions passives canoniques. Le sous-corpus EN-Brochures, contient également bien moins de constructions passives que le corpus EN-Lois, mais davantage que EN-CAB ; or, il s'appuie explicitement sur des textes issus de EN-Lois, ce qui pourrait expliquer pourquoi il contient davantage de passifs. La diminution du nombre de verbes à la voix passive dans le passage du corpus de Jugements au corpus de Résumés de la CSC est moins évidente.

8.1.2 Fréquence des constructions passives canoniques dans les corpus francophones

Nous passons maintenant aux résultats quantitatifs pour le français. Les Tableau 8.3 et Tableau 8.4 présentent ces fréquences. Comme pour les corpus anglophones, un test de chi-carré a d'abord été réalisé pour examiner les différences entre la proportion de passif par groupe verbal trouvée dans les corpus LEX et PLAIN, à partir des fréquences brutes. Nous trouvons une différence significative entre les corpus LEX et PLAIN ($\chi^2(1, N = 434969) = 10659.4, p\text{-value} < .00001$).

Les résultats révèlent également une différence liée aux genres discursifs entre les deux sous-corpus de textes juridiques, puisqu'il y a deux fois plus de constructions

TABLEAU 8.3 – Fréquence de la voix passive dans le corpus FR-LEX

Sous-corpus	Nombre brut de constructions passives	Constructions passives par million de mots	Nombre de GV conjugués	% de constructions passives par rapport au nombre de GV conjugués
FR-Lois	69171	13606	207703	33,3
FR-CA-Jugements	5736	6207	49678	11,5

passives par million de mots dans le sous-corpus FR-Lois que dans le corpus FR-CA-Jugements. Rapporté au nombre de verbes conjugués, on trouve environ trois fois plus de constructions passives dans le corpus de lois que dans le corpus de jugements. À nouveau, ces premières mesures ne concernent que les constructions passives canoniques, ce qui peut expliquer en partie cette différence, comme nous tâchons de le démontrer dans la section suivante.

TABLEAU 8.4 – Fréquence de la voix passive dans le corpus FR-PLAIN

Sous-corpus	Nombre brut de constructions passives	Constructions passives par million de mots	Nombre de GV conjugués	% de constructions passives par rapport au nombre de GV conjugués
FR-DILA	5938	11949	27107	21,9
FR-Brochures	2438	9122	11784	20,7
FR-CA-Résumés	762	10411	5899	12,9

Concernant les textes du corpus FR-PLAIN, le Tableau 8.4 révèle qu'on trouve à peu près la même proportion de passifs dans les textes appartenant au registre administratif du sous-corpus FR-DILA et dans les guides explicatifs du sous-corpus FR-Brochures. Les résumés des jugements de la CSC présentent une proportion moindre de ces structures passives canoniques, qui ne représentent 13% environ des formes verbales finies dans ce sous-corpus, qui contient par ailleurs légèrement plus de constructions passives canoniques que dans le corpus FR-CA-Jugements (12,9% contre 11,5%).

Dans l'ensemble, ces résultats quantitatifs révèlent que les rédacteur.ices des textes du corpus PLAIN utilisent librement le passif canonique, bien qu'en moindre proportion que dans le corpus LEX. Parmi les textes juridiques, c'est sans conteste le discours législatif, en anglais comme en français, qui emploie le plus fréquemment les passifs canoniques en *be/être*. Dans le corpus PLAIN, le passif canonique est particulièrement saillant dans les Brochures (en français comme en anglais), qui contiennent environ 20% de groupes verbaux finis au passif. Cependant, tous les sous-corpus PLAIN contiennent plus de 10% de formes verbales finies au passif, allant à l'encontre des préconisations qui conseillent d'employer 90% de verbes à la voix active. Ces résultats quantitatifs infirment donc partiellement notre hypothèse

de départ, puisqu'ils suggèrent que la recommandation préconisant d'éviter le passif dans la rédaction en style clair n'est pas toujours suivie pas les rédacteur.ices. Le passif est donc bien utilisé dans les textes des corpus LEX comme PLAIN, d'autant plus que nous n'avons analysé jusque-là que les formes canoniques. Avant de nous pencher sur l'analyse qualitative des formes passives canoniques et non-canoniques, nous nous proposons d'identifier les verbes qui apparaissent le plus souvent dans des constructions passives prototypiques dans les deux langues.

8.1.3 Verbes les plus fréquents dans les passifs canoniques

Les fonctions "Concordance" et "KWIC lemmas" de SketchEngine permettent d'identifier les lemmes verbaux les plus fréquemment employées dans les structures passives canoniques de chacun des sous-corpus. Nous présentons d'abord les résultats pour les textes juridiques en anglais et en français, en comparant les sous-corpus de EN-LEX et de FR-LEX, puis le discours simplifié en comparant les sous-corpus d'EN-PLAIN et de FR-PLAIN.

TABLEAU 8.5 – Verbes les plus fréquemment passivés dans le discours juridique anglophone

(a) EN-Lois			(b) EN-CA-Jugements		
Lemme	Fréquence brute	Fréquence relative (pmm)	Lemme	Fréquence brute	Fréquence relative (pmm)
make	3342	1406,9	make	231	317,6
give	1156	486,6	entitle	180	247,5
amend	1066	354	require	156	214,5
treat	841	328,4	establish	122	167,7
satisfy	780	328,4	find	122	167,7
take	714	300,6	consider	117	160,8
provide	620	261	take	111	152,6
require	602	253,4	say	105	144,4

Les passifs du corpus EN-LEX qui sont représentés dans les Tableaux 8.5a et 8.5b correspondent à des formes de type [lemme = *be* + participe passé]. Dans les deux sous-corpus, le verbe qui entre le plus fréquemment dans une construction passive est *make* (*be made*). Sa fréquence relative reflète la différence générale concernant la fréquence des constructions passives canoniques entre EN-Lois et EN-Jugements, puisqu'il est quatre fois plus fréquent dans le premier que dans le second. Deux autres verbes sont communs aux deux sous-corpus : *take*, à peu près deux fois plus

fréquent dans EN-Lois, et *require*, à peu près aussi fréquent dans les deux corpus. Les autres verbes les plus fréquents, comme *amend* dans EN-Lois et *establish* dans EN-CA-Jugements, semblent plus spécifiques à chacun des sous-corpus. Pour les verbes *make*, *take* et *require*, une étude des contextes est nécessaire pour expliquer pourquoi ces verbes sont communs et fréquents dans les deux sous-corpus, alors que la plupart des verbes fréquemment au passif semblent spécifiques au sous-corpus.

TABLEAU 8.6 – Verbes les plus fréquents dans les passifs canoniques du corpus FR-LEX

(a) Dans FR-Lois			(b) Dans FR-CA-Jugements		
Lemme	Fréquence brute	Fréquence relative (pmm)	Lemme	Fréquence brute	Fréquence relative (pmm)
fixer	2287	489,9	tenir	133	143,9
faire	1364	268,3	établir	125	135,28
tenir	1353	266,1	donner	115	124,5
prendre	1342	264	prendre	102	110,4
établir	1338	263,2	interpréter	100	108,2
mettre	1086	213,6	rendre	82	88,7
effectuer	992	195,1	justifier	75	81,2
accorder	963	189,4	fonder	72	77,9

Les Tableaux 8.6a et 8.6b représentent la fréquence des verbes les plus souvent passivés (dans des structures passives canoniques) dans les corpus spécialisés francophones. A nouveau, on trouve plusieurs verbes communs aux deux sous-corpus, en particulier *tenir*, *prendre* et *établir*, et une analyse des contextes de ces structures est nécessaire pour voir s'ils sont employés dans les mêmes contextes dans les deux genres discursifs.

Les Tableaux 8.7a, 8.7b et 8.7c montrent quant à eux les verbes les plus fréquemment passivés dans les sous-corpus d'EN-PLAIN. On retrouve *make* et *take* dans les sous-corpus de textes procéduraux et administratifs, EN-Brochures et EN-CAB, comme dans les corpus spécialisés, suggérant des utilisations similaires, ce que nous proposons ensuite de vérifier en contexte. Ils ont également en commun *give*. EN-CA-Résumés contient un seul verbe en commun avec EN-Brochures, qui est *use*. Les autres verbes les plus fréquemment employés dans des structures passives canoniques sont spécifiques à chacun des trois sous-corpus.

Dans les Tableaux 8.8a et 8.8b, qui présentent les verbes qui sont le plus souvent au passif dans FR-Brochures et FR-DILA, on remarque trois verbes en commun :

TABLEAU 8.7 – Verbes les plus fréquents dans les passifs canoniques du corpus EN-PLAIN

(a) EN-Brochures			(b) Dans EN-CAB		
Lemme	Fréquence brute	Fréquence relative (pmm)	Lemme	Fréquence brute	Fréquence relative (pmm)
make	137	580,9	make	524	655
give	104	441	call	488	608
take	73	309,5	pay	368	460
provide	68	288,3	entitle	320	400
grant	58	246	treat	290	362,5
require	53	224,7	give	211	263,7
use	53	224,7	take	185	237,5
ask	49	207,8	discriminate	72	231,2

(c) Dans EN-CA-Résumés		
Lemme	Fréquence brute	Fréquence relative (pmm)
charge	29	658,3
hear	24	544,8
allow	22	499,4
mean	17	385,9
base	14	317,8
use	13	295,1
find	12	272,4
consider	12	272,4

prendre, *verser* et *accorder*. Le premier est un des verbes le plus souvent au passif dans FR-Lois, ce qui suggère là encore des emplois similaires. Dans le Tableau 8.8c, les verbes que l'on trouve le plus fréquemment dans des structures passives canoniques sont les équivalents de ceux trouvés dans EN-CA-Résumés et semblent pour la plupart liés à la procédure judiciaire, comme par exemple *accuser*, *arrêter*, *condamner* et *juger*.

Nous passons maintenant à l'analyse en contexte des passifs. Nous étudions d'abord les concordances contenant des passifs dans le corpus LEX (8.2) puis dans le corpus PLAIN (8.3), afin d'explorer les fonctions rhétoriques de ces structures. Pour chacun des corpus, nous proposons une analyse contextuelle des passifs prototypiques dans les deux langues et pour chaque sous-corpus (8.2.1 et 8.2.2 pour

TABLEAU 8.8 – Verbes les plus fréquents dans les passifs canoniques dans FR-PLAIN

(a) FR-Brochures			(b) Dans FR-DILA		
Lemme	Fréquence brute	Fréquence relative (pmm)	Lemme	Fréquence brute	Fréquence relative (pmm)
prendre	68	254,5	verser	230	462,8
verser	57	213,3	prendre	172	346,1
accorder	51	190,8	concerner	152	305,9
prononcer	43	160,9	remettre	143	287,8
interdire	40	149,7	accorder	131	263,6
reconnaître	36	134,7	condamner	95	191,2
demander	35	130,8	prononcer	94	189,2
indemniser	34	127,2	délivrer	94	189,2

(c) Dans FR-CA-Résumés		
Lemme	Fréquence brute	Fréquence relative (pmm)
accuser	38	519,5
tenir	24	327,9
autoriser	20	273,2
condamner	20	273,2
arrêter	18	245,9
juger	17	245,9
déclarer	17	232,3
utiliser	16	218,6

le corpus LEX et 8.3.1 à 8.3.3 pour le corpus PLAIN). Notre but est d'identifier leurs fonctions rhétoriques et d'analyser d'éventuels schémas LG dans lesquels ils s'insèrent. Nous proposons ensuite une analyse des formes passives non-canoniques pour chaque corpus (8.2.3 et 8.3.4).

8.2 Formes et fonctions rhétoriques du passif dans les corpus LEX

Les constructions passives canoniques représentent environ un quart des groupes verbaux finis dans les corpus de textes législatifs EN-Lois et FR-Lois, et environ 10% dans les corpus CA-Jugements. Elles semblent donc jouer un rôle privilégié dans le

style juridique, en particulier législatif. Dans cette section, nous présentons donc les fonctions et schémas LG principaux associés de ce passif canonique, en nous appuyant sur les co-occurrences les plus fréquentes. Pour chaque corpus, nous présentons les types de sujets (acteur institutionnel collectif ou individuel, document juridique, principe juridique abstrait, justiciable, etc), et, le cas échéant, certaines collocations particulièrement représentatives de chaque genre discursif considéré. Nous nous concentrons enfin sur les formes de passif non-canoniques, qui sont également fréquentes, bien que moins facilement détectées par la requête en CQL.

8.2.1 Passifs canoniques et *boilerplates* dans le discours juridique

Avant de nous pencher sur les verbes qui apparaissent le plus souvent dans des structures passives canoniques spécifiques, nous nous intéressons à certaines séquences formulaïques récurrentes, figées ou semi-figées, qui apparaissent dans tous les jugements et textes de lois.

Dans les corpus anglophones, on retrouve le passif dans les formules liminaires des deux genres juridiques en question, comme dans les extraits ci-dessous, qui ouvrent les textes dont ils sont issus :

(161) This Act **may be cited** as the Consumer Rights Act 2015. (EN-Lois)

(162) The judgment of McLachlin J. and Abella, Moldaver, Karakatsanis, Wagner, Gascon and Brown JJ. **was delivered** by Wagner J. (EN-CA-Jugements)

(163) The judgment of the Court **was delivered** orally by Abella J. (EN-CA-Jugements)

D'autres phrases récurrentes apparaissent également au passif non pas en ouverture mais en conclusion des textes, en particulier dans les arrêts, à la fin des jugements de la CSC, au moment où les juges présentent la décision prise suite à leur analyse de l'affaire en appel, après avoir exposé leurs arguments. Les exemples (164) et (165) montrent les faibles variations de ces schémas phraséologiques,

(164) The appeals **should be dismissed**. (EN-CA-Jugements)

(165) The appeal **should be allowed**. (EN-CA-Jugements)

Ces formules rituelles nous rappellent la nature hautement standardisée et conventionnelle du discours juridique, qui pare ces textes d'une aura de solennité et de performativité. La phrase de fermeture annonçant la décision de la Cour Suprême du Canada a son équivalent en français, également au passif :

(166) Arrêt (la juge Côté est dissidente) : Le pourvoi **est rejeté**.²

Enfin, le verbe *amend* est particulièrement fréquent dans le corpus EN-Lois et lié à une séquence formulaïque saillante. Il est également lié à une collocation spécifique aux *statutes*, [Nom de la loi + *is amended as follows*], qui se trouve à la fin des textes et apparaît 379,3 fois pmm, lorsque les législateur.ices indiquent tous les textes législatifs et articles de lois modifiés par le texte en question.

(167) The Civil Partnership Act 2004 **is amended as follows**. Omit section 6 and (prohibition on use of religious premises for registration of civil partnership).

(168) Article 56 (alignment with rest of UK) **is amended as follows**.

Ces collocations, qui servent principalement une fonction textuelle et métalinguistique, se retrouvent également dans le discours législatif francophone, sous la forme *les mots (...) sont remplacés par les mots (...)*, comme on le voit ci-dessous :

(169) En Nouvelle-Calédonie, en Polynésie française et dans les îles Wallis et Futuna, les mots : "juge du tribunal d'instance" **sont remplacés** par les mots : "président du tribunal de première instance ou le juge délégué par lui", et le mot : "juge" est remplacé par les mots : "président du tribunal de première instance ou le juge délégué par lui". (FR-Lois)

(170) Aux articles 11, 17, 18-2, 19, 19-2, 20, 23, 29-1A, 29-1B, 29-1, 29-2, 29-4, 30, 34 et 42, les mots : "tribunal de grande instance" sont remplacés par les mots : "tribunal de première instance". (FR-Lois)

Toutes ces collocations ou phrases sont des exemples de *boilerplates*, c'est-à-dire, comme nous l'avons vu au Chapitre 1, des morceaux de textes qui sont copiés et

2. Bien que cette comparaison dépasse le cadre de notre propos, il est intéressant de noter que la Cour de Cassation française utilise quant à elle l'actif pour prononcer ses arrêts : "PAR CES MOTIFS, la Cour : **REJETTE** le pourvoi" suivi cependant d'une expression figée contenant des passifs sans auxiliaire : "Ainsi **fait et jugé** par la Cour de cassation, chambre criminelle, et **prononcé** par le président le vingt octobre deux mille vingt et un." (Décision de la Chambre criminelle de la Cour de Cassation française, 20 octobre 2021, consultée le 5 novembre 2021.)

ré-utilisés d'un texte à l'autre avec très peu (voire pas) de variation. Ici, seul le titre de l'*Act* et le nom du juge qui a rendu le jugement sont modifiés à chaque nouveau texte.

Ces formules figées rappellent la dimension profondément méta-linguistique et auto-référentielle des textes juridiques, associée à leur caractère performatif : au moment où il s'énonce, le texte de loi vient modifier les textes précédents ; au moment où le/la juge le prononce, le jugement vient annuler ou confirmer le jugement des cours inférieures, et ses paroles prennent force de loi en faisant jurisprudence.

On notera que la plupart de ces *boilerplates* contiennent des passifs qu'Huddleston et Pullum (2005) nomment des "passifs courts", c'est-à-dire sans complément en *by*, en particulier dans nos exemples pour l'anglais. Dans l'exemple en français, on voit que la construction contient un GP commençant avec "par", mais il n'exprime pas un complément d'Agent, seulement une mention des mots qui viennent remplacer le texte original, c'est-à-dire un circonstant de moyen ; si, en surface, cette structure ressemble donc à un passif long, en réalité, comme les constructions passives méta-linguistiques vu plus haut, elle ne comporte pas d'Agent. Si l'agent n'est pas exprimé, c'est qu'il est évident à la fois pour les rédacteur.ices et le lectorat (expert aussi bien que non-expert) que la modification émane des institutions législatives ou que la décision est le résultat de l'analyse des juges. Pour les textes législatifs par exemple, il est évident que c'est l'autorité de la loi qui est exprimée. Ainsi le Code pénal français déclare-t-il dans son deuxième article :

La loi détermine les crimes et délits et fixe les peines applicables à leurs auteurs.
(Article 111-2 - Code pénal français)

Cet article liminaire, tout comme dans la formule d'ouverture des textes législatifs du Royaume-Uni (*Be it enacted...*), annonce que tout le reste du texte exprime l'intention et la voix du législateur. Il n'est donc pas nécessaire, ensuite, de préciser pour chaque modification que les textes précédents dont il est question sont amendés *par la loi*. Dans les *boilerplates* issus des jugements, l'agent n'est exprimé que lorsqu'il doit être incarné par un. acteur.ice du droit spécifique et nommé, par exemple le juge Wagner. Ces exemples de constructions passives insérées dans des segments méta-linguistiques se prêtent bien à l'analyse aspectuelle du passif proposée par Beedham (1987). Dans tous ces exemples, c'est le changement d'état du référent du sujet qui est exprimé. Le passif est utilisé pour signaler que c'est le nouvel état du texte (loi ou jugement), résultant à l'action exprimée par le verbe (modification du texte, rejet du pourvoi) qui est au centre de ces *boilerplates*. Le groupe prépositionnel est donc optionnel.

8.2.2 Le passif : un outil pour thématiser les acteurs institutionnels et les objets du droit

Nous passons maintenant à l'illustration qualitative des verbes qui se retrouvent fréquemment dans des structures passives canoniques dans le corpus LEX. Ce que nous nous attachons à démontrer ici est que le passif permet fréquemment de thématiser les acteurs institutionnels et objets du droit (principes, concepts juridiques, textes/documents juridiques) à l'intérieur de schémas LG (Gledhill et al. 2017) à la structure partiellement prévisible. Dans ces schémas lexico-grammaticaux que nous allons décrire en particulier, le passif est employé systématiquement, ce qui soutient l'interprétation du passif comme une forme qui sert une fonction discursive et n'est pas seulement le résultat d'un réagencement syntaxique, contrairement à ce qui est indiqué dans les recommandations pour un style clair.

Les tableaux 8.5a et 8.5b dans la section 8.1.3 montrent que les verbes les plus fréquemment employés dans les constructions passives canoniques dans EN-LEX sont *make*, *take* et *require*. Pour *make*, on trouve dans les deux sous-corpus des structures du type < [Nom prédicatif] + *is made* + [groupe prépositionnel (GP)]>.

- (171) If **an application is made** under this section in respect of a temporary protection order, the Registrar must assign a hearing date, which must be— as soon as practicable. (EN-Lois)
- (172) **An application** for a warrant under subsection **may be made by a constable or the chief executive**.
- (173) **Mr. Groia's mistaken allegations were made** in good faith and were reasonably based. (EN-CA-Jugements)
- (174) None of **the impugned actions or decisions** cited in Mikisew Cree First Nation's application for judicial review underlinewere made by a "federal board, commission or other tribunal". (EN-CA-Jugements)

Dans ces exemples, *make* est le prédicat mais c'est véritablement le COD qui exprime le procès, comme on le voit par exemple avec la structure *make allegations* au passif (173), où c'est le nom *allegations* qui exprime que des accusations ont été portées. Il s'agit donc de prédicats complexes contenant des noms prédicatifs, souvent citées comme caractéristiques du discours juridique (et spécialisé en général). Dans les exemples (174) et (172) tirés de EN-Lois ci-dessus, le sujet est sous la forme d'un GN, *an application*, qui est une nominalisation du verbe *apply* et fait référence à un document juridique. Un autre nom que l'on trouve souvent comme sujet de *be made* dans ce sous-corpus est *order*, qui est l'équivalent de *ordonnance* en français.

On trouve par ailleurs souvent le circonstant métatextuel post-verbal *under this (sub)section*, qui précise que le document juridique en question doit faire référence à ce passage du texte de loi précisément. Le schéma lexico-grammatical est donc de la forme <An order/application (+ auxiliaire) + *be* + *made* + *under this section*>. Dans EN-CA-Jugements, les exemples contiennent également des constructions VN, comme *make a decision* ou *make allegations* comme nous l'avons vu, où le sujet est souvent plus abstrait : *decisions* et *allegations* renvoient on pas à des documents mais à des actions abstraites liées à des procédures ou concepts judiciaires.

Les formes passives contenant *take* qles plus fréquentes dans les deux sous-corpus incluent souvent le prédicat complexe lexicalisé *take into account*. Dans cette construction, c'est le complément prépositionnel *into account* du verbe *take* qui a une fonction prédicative ; l'emploi de cette structure au passif permet ainsi de thématiser les facteurs que les magistrat.es doivent prendre en compte dans l'interprétation de la loi pour rendre une décision, comme on le voit dans les exemples suivants :

- (175) For the purposes of subsection X, the following matters **may be taken into account** in deciding whether a failure, default, or contravention is serious : the amount of money involved ; whether it comprises a single instance or a series of instances ; if it comprises a series of instances,– how many instances it comprises. (EN-Lois)
- (176) As I will explain, the fact that the behaviour occurs in a courtroom is an important contextual factor that **must be taken into account** when evaluating whether that behaviour amounted to professional misconduct. (EN-CA-Jugements)

Dans les deux cas, le référent du sujet est le plus souvent un des facteurs qui peuvent (175) ou doivent (176) être pris en compte pour analyser une infraction. On note cependant des différences d'emploi de *take* entre les deux genres discursifs. Dans les textes législatifs, il est notamment utilisé au passif pour thématiser des éléments de la procédure en question, avec un sens abstrait (étapes à suivre) ou concret (empreintes digitales à prendre) :

- (177) The steps to **be taken** under subsection above shall include giving the requisite information both orally and in writing.
- (178) Fingerprints **may be taken** from the detained person only if they **are taken** by a constable– with the appropriate consent given in writing, or without that consent under sub-paragraph 2.

Ces passifs avec *take* servent donc parfois à thématiser des sujets abstraits et à guider les agent.es de l'exécutif, notamment en indiquant les informations qu'ils

doivent donner à ou obtenir des justiciables (en particulier le consentement). Dans EN-CA-Jugements, le schéma LG < *be taken as* + V + ing > est davantage métatextuel et peut être compris comme un synonyme de *understood* :

- (179) The phrase "obviously relevant" **should not be taken as** indicating a new standard or degree of relevance : Jackson, at para 125, per Watt. Rather, this phrase simply describes information that is not within the investigative file, but that would nonetheless **be required to be disclosed** under Stinchcombe because it relates to the accused's ability to meet the Crown's case, raise a defence, or otherwise consider the conduct of the defence. (...) These qualifiers are significant, as they contemplate that not all police records will be subject to first party disclosure. (EN-CA-Jugements)
- (180) In saying this, I do not wish to **be taken as** suggesting that the categorical approach established in this Court's jurisprudence is without difficulties. (EN-CA-Jugements)

Dans le premier cas (179), le passif sert à thématiser un extrait de la jurisprudence qui est cité : *the phrase "obviously relevant"* que le/la juge commente et explicite dans le cadre de son analyse afin de préciser le sens du principe juridique qui s'applique à l'affaire jugée. La phrase pourrait être glosée par "*The phrase "obviously relevant" is not to be understood as...*", et signale un commentaire de cette expression par le/la juge, qui restreint son sens (comme le montre l'adverbe *simply*) et indiquer quelles informations peuvent être transmises ou non à la police. Dans le second cas (180), la structure passive est utilisée avec un pronom de première personne, pour que le ou la juge reformule ou précise ce qu'elle veut ou ne veut pas dire. Le ou la juge exprime sa position (*stance*) et s'assure, grâce à l'emploi d'une structure passive, qu'il ou elle a été bien compris.e par ses confrères et consœurs. Ces exemples font écho à l'utilisation de ces constructions passives dans EN-Lois, mais ces collocations au passif semblent avoir dans les motifs de jugement à la fois une valeur argumentative et interpersonnelle, là où elles ont plutôt une valeur descriptive et procédurale dans EN-Lois, puisqu'elles visent à guider les magistrat.es dans leur analyse.

Le verbe *Require* apparaît également dans des constructions passives canoniques fréquentes dans les deux genres discursifs :

- (181) This section applies if a hearing **is required** or held (for example, because the court is notified, under section 147 or 151, that the respondent, an associated respondent, or a person to whom section 151 applies, wishes to be heard). (EN-Lois)

- (182) Where a person **is required** to reside in an accommodation centre by virtue of subsection the Secretary of State must arrange for the provision of accommodation for the person in an accommodation centre. (EN-Lois)
- (183) Even when an audit reveals the emergence of pay inequity during the previous five years, only adjustment payments going forward are payable. (...) Employers **are required** to post the results of the audit, but not the date on which pay inequity emerged, thus obscuring when adjustment payments ought to have been made. (EN-CA-Jugements)
- (184) Therefore, correctness review **is required** to ensure proper respect for the judiciary's constitutionally guaranteed place in our democracy. (EN-CA-Jugements)

Dans les extraits (181) et (182) d'EN-Lois, on note que le segment *is required* apparaît dans des propositions conditionnelles en *if* ou *where* et a pour sujets des GN qui font références à des éléments de procédure abstraits (*hearing*), ou à des animés humains abstraits référant aux justiciables (*person*). Dans les motifs de jugements, *require* est utilisé pour faire un rappel de la loi qui s'applique à un sujet collectif, qui fait ici référence aux employeurs, pour commenter de possibles contournements de la loi (183). Dans l'exemple (184), c'est un terme, *correctness review* qui est sujet de *required* : il correspond à l'un des principes juridiques fondamentaux de la révision judiciaire dans le cadre du droit administratif. Le juge indique dans cet exemple quelle approche la Cour devrait adopter envers la décision des cours inférieures pour juger l'affaire en question en respectant la constitution³.

D'autres lemmes verbaux sont plus spécifiques à chacun des genres discursifs en question. Nous présentons brièvement quelques exemples, afin d'identifier quels types de sujets sont thématiques et quelles constructions passives entrent dans des collocations spécifiques.

- (185) Any statement made in court to a Judge or a witness by a person providing communication assistance **must**, if known by the person making that statement to be false and intended by that person to be misleading, **be treated as perjury** for the purposes of sections 108 and 109 of the Crimes Act 1961. (EN-Lois)
- (186) The Tribunal **must be given access** to classified information that— was relied on to make a decision that is on appeal to the Tribunal; or is first raised in the course of an appeal to, or a matter before, the Tribunal. (EN-Lois)

L'exemple (185) présente un exemple d'un schéma LGde la forme <[GN = document / élément de la procédure] + is to be/must be treated as + [GN] + for the

3. Il existe deux standards pour la révision judiciaire : *correctness review*, en français "norme de la décision correcte" et *reasonableness review*, "norme de la décision raisonnable". (Dunsmuir v New Brunswick, [2008] 1 SCR 190)

purposes of + [GN = référence à un article de loi] >. La présence au sein de ce schéma de la préposition complexe *for the purposes of this section*, segment extrêmement fréquent dans les textes de lois anglophones, confère à ce schéma lexico-grammatical une dimension méta-textuelle. Concernant l'exemple (186), il illustre cette fois la thématization d'un acteur institutionnel collectif (*tribunal*), suivi de <(auxiliaire de modalité) + *be given* + *access / notice / information*. Le GN après *given* est relativement variable, puisqu'il peut s'agir de *access* ou *information*, mais la collocation fait toujours référence à l'information qui est donnée aux tribunaux au cours de la procédure judiciaire. Ces structures servent à nouveau à guider les magistrat.es et agent.es dans l'exécution du texte, en précisant les conditions d'application de la loi, qui peuvent changer selon les situations individuelles (c'est pourquoi ces phrases contiennent des propositions conditionnelles ou concessives). Surtout, ces passifs apparaissent dans les deux cas dans une structure contenant l'auxiliaire de modalité *must*. Le modal a une valeur déontique : le contenu de la proposition relève de l'obligation envers l'institution (*the Tribunal*) ou son représentant (*the judge*), ce qui est une caractéristique du discours législatif, qui est, nous l'avons vu, fondamentalement normatif.

Dans EN-CA-Jugements, les passifs canoniques les plus fréquents, comme *entitled* et *found*, permettent également de thématiser les acteurs institutionnels collectifs (comme la Cour) ou les parties (ci-dessous, l'accusée).

- (187) Justice Moldaver states that it was "not reasonably open" to the Appeal Panel to find that Mr. Groia's allegations lacked a reasonable factual basis : M.R., at para 134.(...) With respect, the Appeal Panel **was entitled** to make the findings of fact it made. (EN-CA-Jugements)
- (188) If, as in the case at bar, an element of the *actus reus* is missing at the time of the alleged offence, the defendant **cannot be found guilty**. (EN-CA-Jugements)

Un co-occurent spécifique de *made* fait apparaître une structure au passif particulièrement intéressante qui semble être spécifique au genre discursif du jugement, et est équivalente à *établir* dans le corpus francophone :

- (189) Both the *actus reus* and the *mens rea* **must be made out**. While there is no doubt that Mr. Carson had a guilty mind, establishing the *mens rea* is insufficient, in and of itself, to **make out** the offence.
- (190) The Crown is arguing that the elements of the offence **have been made out**, not that the accused should have received a longer sentence or period of parole ineligibility.

Ici, il s'agit d'un emploi spécialisé du verbe à particule *make out*⁴, et non plus simplement du verbe *make*. Le schéma LG est de la forme < [Principe juridique] (+ auxiliaire modal/d'aspect) + *be* + *made out* >. L'auxiliaire *must* dans le premier exemple a là aussi une valeur déontique : le juge exprime une obligation, rappelant ce que la loi prévoit : pour qu'une infraction criminelle soit établie, il faut à la fois que l'élément matériel (*actus reus*), c'est-à-dire l'acte criminel (par exemple, le meurtre), et l'intention (*mens rea*) soient démontrées.

L'équivalent français du *make out* anglophone est *être établi*, comme nous le voyons dans ces deux extraits de jugements :

(191) L'*actus reus* et la *mens rea* doivent tous deux **être établis**. (FR-CA-Jugements)

(192) La juge Schaffter a conclu que l'atteinte à l'art 7 **avait été établie**. (FR-CA-Jugements)

Comme pour l'anglais, ce sont des principes juridiques qui sont thématiques, dans la proposition principale ou dans une proposition subordonnée rapportant les conclusions du juge de la cour inférieure. A nouveau, l'interprétation aspectuelle du passif de Beedham permet une analyse élégante du recours au passif dans ces extraits. En effet, l'absence de complément en BY n'a pas pour but d'occulter l'agent, mais permet surtout de focaliser l'état résultant visé ou obtenu. Le passif permet par exemple d'affirmer la nécessité que le principe juridique soit établi dans l'affaire concernée. Dans ce contexte, il n'est pas nécessaire de mentionner l'agent, puisqu'il est évident que c'est par exemple la cour d'appel ou le ministère public qui doit établir l'infraction.

Etablir est également souvent passivé dans FR-Lois, pour thématiser là encore un principe juridique, par exemple *la filiation* dans l'exemple (193). Une autre séquence particulièrement fréquente est liée au verbe *fixer*, et illustrée dans l'exemple (194).

(193) Lorsque la filiation **est établie** à l'égard de l'un des parents plus d'un an après la naissance de l'enfant, l'autre parent exerce seul cette autorité. (FR-Lois)

(194) Les conditions d'application du présent article **sont fixées** par décret en Conseil d'Etat. (FR-Lois)

4. Bien loin de ces emplois généraux, et illustrant la polysémie externe de la terminologie juridique.

Le premier exemple montre un emploi équivalent aux jugements. La seconde collocation présentée est particulièrement productive dans FR-Lois, avec 450 occurrences pmm de la séquence semi-figée <[Principe juridique] *est fixée(e) par décret*>. Le principe en question peut être abstrait, comme les conditions d'application de la loi, mais il peut s'agir également d'éléments plus concrets comme une durée ou des éléments à prendre en compte dans le calcul d'indemnités. Le statut semi-figé de cette séquence, et sa fréquence d'apparition dans les corpus suggèrent que le passif, en tout cas avec ce verbe-pivot, fait véritablement partie de la phraséologie du discours législatif français.

On trouve également des prédicats complexes qui sont communs aux deux sous-corpus de FR-LEX. Par exemple, *prendre* est également souvent employé sous la forme de l'expression lexicalisée *pris(es) en compte*, qui a une fréquence de 101,9 pmm dans FR-Lois et 39,0 pmm dans FR-CA-Jugements. Ces emplois sont illustrés dans les exemples ci-dessous.

- (195) Pour la détermination du montant de l'allocation, sont **prises en compte** les rémunérations soumises à contribution et correspondant à ces périodes d'affiliation. (FR-Lois)
- (196) En conséquence, les actes de violence dont M. Suter a été victime aux mains de justiciers n'agissant pas au nom de l'État **peuvent être pris en compte** au moment de déterminer la peine appropriée. (FR-CA-Jugements)
- (197) Enfin, deux éléments contextuels externes — qui relèvent tous deux de l'expertise et des compétences de la Commission — **doivent être pris en compte**. D'abord, la Commission a adopté le par. 26.2(1) dans sa version actuelle par suite de l'inquiétude suscitée par l'augmentation du nombre de décès au travail dans le secteur forestier (...) (FR-CA-Jugements)

Les sujets de ces constructions passives font référence à des concepts administratifs abstraits dans les textes législatifs français (195), en particulier liées à l'argent ou à une durée (*rémunérations*). Dans FR-CA-Jugements, il font davantage référence à des éléments concrets liés aux faits de l'affaire, comme des actes de violence (196), mais qui peuvent aussi présenter un plus haut degré d'abstraction, en particulier lorsqu'ils servent d'argument juridique par exemple dans l'analyse et la présentation des motifs par les juges. Deux de ces exemples permettent également de souligner le rôle sémantique du nom, puisque dans les deux premiers exemples, les sujets *rémunérations* et *actes de violence* sont eux-mêmes des noms prédictifs qui expriment un procès ; on peut donc parler de métaphore grammaticale imbriquée, et on a ici deux beaux exemples d'*intricacy*, avec plusieurs noms prédictifs, un prédicat complexe et un passif dans la même proposition.

Dans les exemples de FR-CA-Jugements ci-dessus, on note l'emploi des auxiliaires modaux *devoir* et *pouvoir*, tout comme de nombreux exemples du corpus anglais contiennent les modaux *must* et *may* ou *can*. On trouve donc fréquemment dans le discours juridique des marqueurs de modalité associés à des constructions passives, comme nous l'avons vu plus haut et comme évoqué par Maley (1994).

Enfin, en français, dans les deux sous-corpus, le passif est utilisé dans la collocation <[Acteur institutionnel collectif] *est tenu de* + proposition> pour rappeler les obligations des acteurs institutionnels collectifs non-humains, comme le ministère public ou les institutions administratives :

- (198) Le ministère public **était tenu** d'établir une forte apparence de droit quant à l'existence d'un outrage criminel au tribunal. (FR-CA-Jugements)
- (199) Le droit de toute personne à l'information est précisé et garanti par le présent chapitre en ce qui concerne la liberté d'accès aux règles de droit applicables aux citoyens. Les autorités administratives **sont tenues** d'organiser un accès simple aux règles de droit qu'elles édictent. (FR-Lois)

Les verbes les plus fréquemment passivés dans les corpus EN-LEX et FR-LEX, quel que soit le genre discursif, confirment que les prédicats complexes sont caractéristiques du discours juridique, et servent à thématiser les acteurs du droit (Cour, institutions, parties) ainsi que les principes et éléments de la procédure (concrets et abstraits) ou encore les documents juridiques. Les passifs les plus fréquents sont insérés dans des collocations récurrentes présentant des restrictions sémantiques (réfèrent du sujet), et syntaxiques (contraintes sur le verbe-pivot au passif et les GP). Ces analyses semblent confirmer que les passifs canoniques dans ces corpus sont liées à des schémas LG, qui sont cohérents avec les résultats obtenus dans des études précédentes sur le discours juridique (Gledhill et al. 2017). Nous nous tournons maintenant vers les formes non-canoniques de passif dans ce corpus.

8.2.3 Passifs sans auxiliaires et cumul de formes passives dans le corpus LEX

En français comme en anglais, les résultats quantitatifs révèlent que les sous-corpus CA-Jugements contiennent moins de constructions passives canoniques.

On peut d'abord se demander si le corpus EN-CA-Jugements contient peut-être davantage de verbes qui dénotent des procès qui ne sont pas passivables. C'est le

cas par exemple des procès mentaux qui permettent d'exprimer le raisonnement des juges de la CSC, tels que *believe*, *conclude*, *explain* ou *find* ci-dessous.

(200) As a result, I **would find** that 26.2 is *ultra vires*, and I would allow the appeal on that basis.

Souvent suivis de propositions complétives, ces verbes ne sont pas ou difficilement passivables (*It was found ?by me that...*). La tendance des juges à utiliser ces verbes à la première personne pour exprimer leurs décisions sert une fonction rhétorique argumentative, qui pourrait expliquer qu'il y a moins de passifs canoniques dans les sous-corpus jugements en anglais et en français.

Cependant, on peut également faire l'hypothèse que cette différence serait due au fait que les textes du corpus "CA-Jugements" utilisent moins de structures passives canoniques, et davantage de passifs sans auxiliaire (*bare passives*). Ces formes non-canoniques de passives, dont la présence est attestée dans les textes de jugements, ne sont pas détectées par la recherche CQL des formes finies de passif.

En effet, il suffit de parcourir quelques extraits des motifs de jugements en anglais pour s'apercevoir que les passifs sans auxiliaire y sont légion.

- (201) More importantly, in reading the LSA as a whole, it becomes readily apparent that the functions, duties and powers **set out** therein relate only to the governance of the LSUC itself, to the provision of legal services by lawyers, law firms and lawyers of other jurisdictions, and to the regulation of articulated students and licensing applicants. The LSUC's functions, duties and powers **are**, in short, **limited** to regulating the provision of legal services, starting at (but not before) the licensing process – that is, starting at the doorway to the profession. The LSUC's by-law making authority is similarly **constrained**. Each of the matters **listed** in 62 ("By-laws"), and 62 **read** as a whole, grant the LSUC by-law making powers only for matters relating to the affairs of the Society, and the governing of licensees, the provision of legal services, law firms, and applicants. (EN-CA-Jugements)
- (202) Further, as indicated, despite the criticisms **levelled** at Mr. Groia by Campbell J. and Rosenberg J.A. for the uncivil way in which he had made his allegations against Mr. Naster, the trial judge never once castigated Mr. Groia for the tone or manner of his submissions or the language **used** by him. (EN-CA-Jugements)

Les extraits ci-dessus contiennent un nombre important de passifs, quelques uns canoniques, mais surtout sans auxiliaire. Si, à première vue, il semble que les jugements contiennent moins de passifs, en réalité, il n'en est rien : ils sont en fait présents sous forme de *bare passives*. Ces passifs sont utilisés par les juges pour

intégrer à leur argumentation certains faits de l'affaire jugée et éléments cités, remplissant donc en partie une fonction intertextuelle, comme on le voit avec *Each of the matters **listed** in 62 ("By-laws"), and 62 **read** as a whole* dans l'exemple (201).

Ces passifs non-canoniques ajoutent à la complexité nominale, en post-modifiant des GN composés de plusieurs noms, par exemple *functions, duties and powers*, ainsi que des GN dont le noyau est postmodifié non seulement par des GP en *by* mais aussi d'autres GP comme *for the uncivil way in which he had made his allegations against Mr. Naster*. Ils participent donc de la cohésion discursive, en thématissant les éléments mis en avant par les juges. De plus, ces textes ayant une forte dimension nominale (comme nous l'avons vu au Chapitre 6), ces passifs sans auxiliaire renforcent la cohésion discursive et contribuent au style nominal de ces textes. On trouve ces mêmes types de passifs dans le corpus francophone équivalent, comme on le voit dans cet extrait tiré de la traduction des motifs de jugements de la même affaire :

- (203) La décision de ne pas agréer cette faculté de droit en raison des restrictions **prévues** dans le Covenant — un code de conduite **protégé** par la législation provinciale relative aux droits de la personne — constitue une atteinte profonde à la liberté de religion et est contraire au devoir de neutralité religieuse de l'État. De plus, même si l'"intérêt public" devait être interprété largement, comme le prétend la LSBC, la reconnaissance de la faculté de droit **proposée** par TWU ne serait pas incompatible avec l'intérêt public ainsi **interprété**. (FR-CA-Jugements)

Les passifs sans auxiliaire se trouvent ici dans des GN enchâssés, comme "*protégé par*", qui est suivi d'un syntagme prépositionnel contenant un GN post-modifié par un groupe adjectival (relative à + GN). Dans la phrase suivante, on trouve à la fois un passif canonique et deux GN dont les noyaux sont post-modifiés par des passifs sans auxiliaire, que l'on pourrait gloser par des relatives contenant des passifs canoniques. Ces formes de passifs participent donc de la complexité syntaxique au niveau du GN. L'exemple (197) ci-dessus montre donc qu'en français aussi le passif sans auxiliaire est utilisé pour post-modifier des GN qui font référence à des concepts juridiques abstraits ou des textes juridiques. Si, à première vue, ces sous-corpus semblent contenir moins de passifs, c'est simplement qu'ils sont présents sous forme non-canonique.

Les sous-corpus législatifs ne sont pas non plus exempt de passifs sans auxiliaire, qui apparaissent souvent conjointement avec des passifs canoniques. Tous les corpus juridiques contiennent donc de nombreux exemples de passages ou phrases qui cu-

mulent différentes formes de passif. Dans EN-lois, on trouve en particulier le cumul de passifs canoniques et sans auxiliaire :

- (204) If fingerprints **are required** by section 18 to **be destroyed**, any copies of the fingerprints **held** by the law enforcement authority **concerned must also be destroyed**.

Le passif sans auxiliaire se trouve donc employé pour thématiser un élément de la procédure policière, les empreintes digitales, et le document juridique qui en garde la trace. Tous deux sont suivis d'un complément d'agent : le GN modifié est rendu plus complexe par la présence de ce groupe verbal post-modifié par un GP. Cette phrase est complexe structurellement, car elle contient des cascades de noms : comme dans les extraits de jugements analysés plus hauts, les passifs sans auxiliaire participent du style nominal des textes législatifs. On comprend donc l'argument des défenseur.euse.s du style clair en voyant ces exemples. Le risque ici n'est pas l'incomplétude de l'information (puisque l'agent est mentionné), mais d'éventuelles difficultés à comprendre la structure de la phrase, qui demande un effort cognitif pour extraire toute l'information contenue dans ces GN complexes.

Dans le corpus FR-LEX, on trouve également une assez grande variété morphologique de passifs, cumulée à des passifs canoniques.

- (205) Lorsque l'enregistrement ne **peut être effectué** en raison d'une impossibilité technique, **il en est fait mention** dans le procès-verbal d'audition qui précise la nature de cette impossibilité. Le procureur de la République en **est immédiatement avisé**. (FR-Lois)
- (206) Lorsque le prévenu est en détention provisoire, le jugement au fond **doit être rendu** dans les deux mois qui suivent le jour de sa première comparution devant le tribunal. Faute de décision au fond à l'expiration de ce délai, **il est mis fin** à la détention provisoire. Le prévenu, s'il n'est pas détenu pour une autre cause, **est mis** d'office en liberté. (FR-Lois)

Ces exemples suggèrent que le passif canonique est associé à des formes de passifs impersonnels apparaissant dans des expressions lexicalisées, *il (en) est fait mention* et *il est mis fin*. Dans les deux sous-corpus francophones, on trouve par ailleurs des passifs morphologiquement actifs, correspondant à des passifs pronominaux décrits par Rouski (2015).

- (207) Les tiers ne **peuvent** ni demander l'exécution du contrat ni **se voir contraints** de l'exécuter, sous réserve des dispositions de la présente section et de celles du chapitre III du titre IV. (FR-Lois)

- (208) Lombard invoque la clause d'exclusion de garde, direction ou gestion des polices d'assurance qui **se lit** comme suit : **Sont exclus** de la présente assurance : . . . La privation de jouissance, la détérioration ou la destruction : . . . De biens meubles dont vous avez la garde ou sur lesquels vous avez pouvoir de direction ou gestion (...) (FR-CA-Jugements)

Dans l'extrait (207) de FR-Lois, le passif pronominal *se voir contraints* peut être glosé par *être contraints*. Dans l'exemple de jugement (208), *se lit* est un ergatif, c'est-à-dire que le verbe a une morphologie active et pronominale, comme si l'action de lire provenait de la proposition elle-même, mais le sujet ne peut être analysé comme faisant référence à l'Agent; on pourrait gloser cette occurrence par *doit/peut être lue comme suit*. Le sujet du verbe est un extrait de loi et cette construction introduit du discours direct, la citation d'un article de loi, lui-même dans une construction passive canonique mais qui présente une inversion de la syntaxe canonique sujet-verbe : *sont exclus de la présente assurance + sujet*. On voit donc ici un nouvel exemple de cumul de formes passives canoniques et non-canoniques. Le sujet du verbe consiste en une longue énumération de GN, caractéristique, comme nous l'avons vu, du discours législatif; il ne pourrait donc pas être placé en début de phrase car le groupe verbal serait alors tout à la fin de la phrase, ce qui serait trop lourd (*top heavy*). C'est donc pour une raison de présentation de l'information qu'il y a une inversion de la syntaxe canonique : l'inversion sert à signaler le début de cette énumération. Dans ces exemples comme dans la plupart des exemples précédents, l'agent n'est pas exprimé, ce qui confirme que c'est la thématization des éléments sujets, dans le cas présent, les parties (*les tiers*) qui importe, et qui est permise par le passif. Cela confirme l'analyse de Rouski (2015), qui voit le passif comme une stratégie discursive qui permet de présenter les résultats des actes de prescription de l'autorité. Cette interprétation est donc assez proche de l'interprétation aspectuelle du passif présentée par Beedham (1981).

Que pouvons nous conclure sur l'utilisation du passif dans EN-LEX et FR-LEX ? Le passif, sous ses formes canoniques et non-canoniques, apparaît donc comme participant d'une "stratégie extensive de thématization" des concepts, documents et participant.es du domaine du droit, selon l'analyse de Banks (2008). Il sert ainsi à thématiser les différents acteurs et objets du droit : principes juridiques, documents et textes de loi, acteurs institutionnels (police, tribunal) et justiciables désignés par un nom générique (*a person, le prévenu*).

Dans les textes législatifs, le passif semble jouer un rôle clé dans l'expression des conditions d'application et d'interprétation de la loi, tandis qu'il appuie la démarche argumentative des juges dans le discours judiciaire. L'utilisation de formes

non-canoniques à l'intérieur de GN permet de maintenir le flux de l'écriture des textes de lois et des jugements, en post-modifiant les constituants de la phrase. Il semble que les passifs sans auxiliaire (*bare passives*) en particulier soient utilisés comme stratégie de cohésion discursive et de présentation de l'information, afin de "compacter" (*pack*) l'information dans des propositions qui contiennent de nombreux GN post-modifiés. En cela, ces formes de passifs ajoutent de la densité lexicale aux textes, et confirment les résultats du Chapitre 6 sur la complexité nominale.

8.3 Usage et fonctions du passif dans le corpus PLAIN

Nous passons maintenant à l'étude de concordances du passif dans le corpus PLAIN. Nous nous intéressons d'abord aux verbes fréquemment employés au passif dans aux sous-corpus EN-Brochures et EN-CAB, FR-Brochures et FR-DILA, dont les textes remplissent la même fonction : guider les usagers à travers les procédures administratives et juridiques. Nous passons ensuite aux sous-corpus EN- et FR-CA-Résumés, qui ne constituent pas le même genre discursif. Puis nous interrogeons les différentes fonctions des passifs dans le corpus PLAIN.

8.3.1 Dans les genres administratifs, passif et thématization des usager.es du droit et de leur expérience

Les corpus de textes administratifs, EN-, FR-Brochures, EN-CAB et FR-DILA, qui ont pour fonction d'informer et de guider les justiciables vis-à-vis de leurs droits, contiennent un plus grand nombre de passifs que les résumés de jugements, en français comme en anglais.

En anglais, les verbes les plus fréquemment passivés dans ces corpus incluent *give*, *take*, *require*, *entitle* mais surtout *make*, que l'on retrouve dans des prédicats complexes identifiés dans le corpus EN-LEX, par exemple ci-dessous *make an order*.

(209) If a bankruptcy restrictions order (BRO) **is made** against you, this rule will also apply as long as the BRO is in force. (EN-CAB)

(210) You **will be legally divorced** 1 month after the date the Dissolution Order **is made**. (EN-Brochures)

Comme dans les textes spécialisés, le prédicat complexe est utilisé pour faire référence au document qui résulte de l'ordonnance, le nom prädicatif exprimant à la fois le procès *order* et représentant le document administratif envoyé par le tribunal. Souvent, dans ces textes administratifs, le sujet du passif est un élément abstrait relatif à une aide ou allocation, et, après le verbe, on trouve un circonstant qui explique comment et à quelle condition cette aide est versée, comme illustré dans les exemples (211) et (212).

- (211) Usually, SSP [Statutory Sick Pay] **is paid** for the first 28 weeks of sickness if you work for an employer. (EN-CAB)
- (212) L'allocation de soutien familial **est versée** sous certaines conditions, selon les règles de droit commun, jusqu'aux 20 ans de l'enfant. (FR-Brochures)

Les rédacteur.ices topicalisent également des principes juridiques et éléments de la procédure judiciaire et administrative à l'aide de passifs canoniques :

- (213) When evidence **is given**, it is possible for the court to be cleared of everyone except the defendant, their lawyers, interpreters and one news reporter.
- (214) Disciplinary action **may be taken** by the employer against workers who breach the confidentiality and data protection policies. (EN-Brochures)
- (215) Le sursis **est prononcé** en même temps que la peine. (FR-DILA)
- (216) La décision **est prononcée** publiquement à la fin de l'audience ou lors d'une audience ultérieure. (FR-Brochures)

Dans ce cas, le passif a pour sujet un nom générique référant à des principes juridiques et éléments de procédure : le passif permet ainsi de donner des explications sur certains concepts à travers une généralisation, exprimant ainsi l'universalité du droit. Parfois, cette universalité est exprimée en mettant en position thématique un référent humain générique, (comme dans les textes juridiques), par exemple (*personne* et *condamné* dans l'exemple (217) :

- (217) Si la personne **condamnée** a plus de 70 ans, elle peut **être libéré** quelle que soit la durée de la peine lui restant à effectuer. La libération conditionnelle lui **est accordée** dès lors que sa réinsertion est assurée. Le condamné doit justifier qu'il sera **pris en charge** à sa sortie ou qu'il bénéficie d'un hébergement.

On note que dans ces exemples, l'Agent n'est pas exprimé car il est soit évident, soit non nécessaire, puisqu'ici c'est le sort de la personne condamnée avant et après sa libération conditionnelle qui est au centre du discours. Lorsque l'Agent est précisé, c'est par exemple lorsque plusieurs acteur.ices du droit peuvent intervenir dans une procédure mais que cet intervenant change selon les contextes :

- (218) La libération conditionnelle est accordée : **par le juge de l'application des peines** pour les condamnés à jusqu'à 10 ans de prison, **par le tribunal de l'application des peines** pour les autres condamnés.

Les textes contiennent donc des passifs dont les sujets font référence à des principes juridiques abstraits, comme dans EN-LEX et FR-LEX. Cependant, une étude des contextes dans ces sous-corpus montre aussi que les constructions passives canoniques en collocation avec des pronoms de deuxième personne sont sur-représentées. Cela signifie que dans les propositions contenant un verbe au passif, c'est *you* ou *vous* en français, ou un GN commençant par *your/votre*, qui en est très fréquemment le sujet. Cet emploi de la deuxième personne correspond au style conversationnel et interactif correspondant à la *conversationnalisation* du discours institutionnel (Turnbull 2018). Les institutions et les associations s'adressent directement au justiciable, comme dans un dialogue imaginaire, conformément aux recommandations pour rédiger en style clair. C'est donc soit les justiciables eux-même qui se trouvent en position thématique sous forme de pronom de deuxième personne (*you*), soit un élément de la procédure juridique précédé par un déterminant possessif de deuxième personne, ou les deux :

- (219) **You're unlikely to be given bail** if : **you are charged** with a serious offence, eg armed robbery ; **you've been convicted** of a serious crime in the past ; **you've been given bail** in the past and not stuck to the terms (EN-Brochures)
- (220) **The amount you owe** to your creditor **can only be taken** out of the money you earn above this amount. (EN-CAB)
- (221) Your circumstances can **be checked** at any time while you are claiming Disablement Benefit. Benefit fraud is a criminal offence and **you can be prosecuted** or **asked** to pay a penalty. (EN-CAB)

Ces pronoms sont liés à la ré-orientation du discours opérée par les rédacteur.ices des textes de diffusion du droit, qui reformulent le discours juridique pour placer les justiciables et leur expérience du droit au centre de leur discours. Le passif est utilisé

ainsi que diverses structures syntaxiques, telles que des subordonnées relatives (*the amount you owe*) ou complétives (*asked to pay a penalty*), qui ajoutent un degré de complexité syntaxique au niveau de la phrase (*intricacy*). Pour le français, on observe également ce mouvement de personnalisation à travers l’emploi de structures passives canoniques et de la deuxième personne, avec le pronom *vous* ou le déterminant *votre*. Dans les exemples suivants, le sujet de la construction passive est à chaque fois un document juridique, mais inclut un pronom de deuxième personne comme complément d’attribution.

- (222) L’aide juridictionnelle **peut vous être accordée** pour tout ou partie du procès et pour faire exécuter une décision. (FR-Brochures)
- (223) Remise de la carte : la carte **vous est remise** par la préfecture ou la sous-préfecture de votre domicile (selon le lieu de dépôt de votre demande). (FR-Brochures)
- (224) À votre arrivée en France, votre carte **vous sera remise** à la préfecture ou la sous-préfecture de votre domicile sur présentation de votre visa. (FR-Brochures)

Dans l’exemple (223), le passif est suivi d’un circonstant en *par*, mais on trouve des exemples où le GP ne commence pas par *par* mais par *à* : *à la préfecture* (224). Cette équivalence entre les prépositions semble à nouveau aller dans le sens de l’interprétation aspectuelle du passif. En effet, on voit ici que le passif est parfois suivi d’autres GP : le complément d’agent en *by* ne serait qu’un groupe propositionnel comme un autre. C’est l’état résultant, (que la carte soit en possession du justiciable après passage en préfecture), ainsi que le circonstant de lieu (la préfecture), qui importent ici, pas l’agent.e qui remet la carte. Le passif canonique est donc utilisé spontanément par les rédacteur.ices pour leur présenter les procédures administratives et judiciaires, en ré-orientant le discours vers des situations relatives à leur propre expérience, notamment en s’adressant directement à eux :

The rewriting enables the reader to see the situation from his experiential point of view, so that abstract knowledge is associated with familiar situations and becomes practical knowledge. (Turnbull 2018)

Concernant l’environnement grammatical de ces passifs canoniques, il sont souvent, en particulier en anglais, associés à la modalité, comme c’était d’ailleurs le cas dans EN-LEX et FR-LEX. Nous prenons l’exemple de l’anglais ; les Figures 8.1 et 8.2 montrent des exemples de concordances de formes de passifs canoniques en collocation avec des modaux.

FIGURE 8.1 – Passifs canoniques et auxiliaires de modalité dans EN-CAB

Left context	KWIC	Right context
... If you have received an overpayment of benefit it does not necessarily mean that you	will be suspected	... or be guilty of benefit fraud if you were unaware of what you were doing. However, the benefit officer
... for example, the Bible or the Koran, in your hand and you'll read from a card (or the card	will be read	... for you) that you will swear that everything you say will be the truth. If you don't have a religious belief
... ay want to use a loss assessor, who can help you prepare the claim and arrange for items to	be valued	... You can find a loss assessor by searching online. Do you have to claim on your insurance
... make sure you have enough cash for day to day expenses because once a bankruptcy order	is made	... your accounts will usually be frozen. If your application is accepted and a bankruptcy order is made
... Part C of your application form at a review of the administration order. A review	can be asked	... for by you, the court or your creditors. It usually happens if you're struggling to make your adminis
... old or dealt with within a certain period, usually three years after the bankruptcy order, it	may be returned	... to you. If you come by any new assets after you've been discharged, these will usually remain yours
... However, you can use another period of time if this is more appropriate. You	will be considered	... to care for the child on a day-to-day basis if you meet the 104-night rule. Even if some care is done
... ily. If you're staying somewhere in the UK for less than three months, you	can ask to be registered	... with a GP on a temporary basis. You can also register temporarily if you move around the country
... Uncertified photocopies are not accepted. A certified copy of a decree absolute	may be obtained	... from the court which decided the divorce. This can take about a week evidence of nationality. If
... ur home and the amount of the mortgage that you already have. You	will not usually be expected	... to remortgage if the new mortgage would extend beyond the existing mortgage term or beyond your state
... credit, though you may still be affected. After six years, the administration order	will be removed	... from the Register and your credit reference file. This means it will no longer affect your credit rating
... des expenses and VAT and get a breakdown of these costs find out what charges, if any,	will be made	... if the sale falls through before contracts are exchanged. Exchange of contracts When contracts are
... insurance is also known as permanent health insurance. The amount of income you	are allowed	... to claim will not replace the exact amount of money you were earning before you had to stop work. This
... ny hours you want to volunteer each week. You won't be paid for your work, but	can be paid	... out of pocket expenses. This includes money spent on food, travel or childcare costs whilst volun
... y not have realised you were exempt from the cap. If the decision is revised, any money you	're owed	can be backdated to the date when the original decision was made. More about changing a benefit
... make monthly repayments. State pension If you're getting your state pension, it	will be taken	... into account when you work out how much you can afford to pay into an IVA. Personal or occupat
... rt at all to give the insurer accurate information, your claim will be refused and the policy	will be cancelled	... Your premiums will normally be refunded to you but may be kept if there is clear evidence of frau
... t have a sign on their premises saying that they're not responsible for any damage - this	can be considered	... an 'unfair term', which means they're still liable - you should tell them this. Claiming insurance Def
... l. If the decision maker agrees that there was an official error a revision decision	will be made	... which changes the decision from the date it was made. If the outcome of the revision is that you w
... information about the person whose file it is. Any information about anyone else	will be blanked	... out. Local authorities have 40 days from the date of your request to send out the files. This

FIGURE 8.2 – Passifs canoniques et auxiliaires de modalité dans EN-Brochures

... additional payments for the purchase of supplementary foods. A one-off payment	may be provided	... to eligible, pregnant women receiving asylum support to help with costs arising from the birth of
... m page 19) is a guide to the patients' rights and pledges contained in the Constitution, and	can be used	... for reference to explain what the Constitution means for patients. It also sets out the re
... fied. Economic loss, and any emotional or other non-economic loss suffered	may also be claimed	... What makes a successful claim? For the claim to succeed, the Tribunal must
... These rules will be explained to you at the time. They may mean that you	can be given	... treatment for your mental disorder even though you do not consent. (Section 3a of the
... nuisance or disturbance. If they refuse to leave without a reasonable excuse, they	may be prosecuted	... Those found guilty of the offence may be liable to a fine of up to £1,000. Where
... rded the benefit of the doubt and treated as children until a professional age assessment has	been completed	... What can I expect when I make my asylum application? You can expect to
... le by the Court, otherwise the complaints will not even be examined. Cases	can only be brought	... to the Court after domestic remedies have been exhausted, in other words, individuals complain
... t to call our FREE helpline: Asylum Support Application UK on 0800 8000 631 This service	can be provided	... in a language you understand. Further information regarding asylum support can be fo
... ment with a service led by a consultant. Information to help you make your choice	can be found	... on the NHS Choices website (www.nhs.uk). CCGs are expected to promote this inform
... tional need. How long will asylum support be provided for? Support	can only be provided	... until a decision is made on your asylum application and whilst an in-country appeal is pending.
... any countries will not extradite their own citizens. Be aware: Abducted children	can not be extradited	... If the country extradites the abductor to Canada, your child may not come with them.
... hysical disabilities can happen by themselves or along with other problems. They	can be caused	... by the same factors as learning disabilities. One specific physical disability is cerebral p
... ies some of the extra rights that victims have. Until a person is convicted, a victim	may be referred	... to as a complainant and the alleged offender will usually be called the accused or the defendan
... oon as possible on 020 8196 4524. If you are vulnerable and are in Scotland you	may be allowed	... to have your screening interview in Glasgow. This however is at the Home Office discr
... y affects your ability to work that taking these steps would not be reasonable. You	may be called	... to see a doctor or nurse who will complete an assessment and send it to the DWP. It w
... e most appropriate treatment is and why. Where required, a written care plan	can also be requested	... Responsibility "Please participate in important public health programmes such as vac
... dition, the independent assessments of service quality from the Care Quality Commission	can be used	... by all involved in planning, influencing and scrutinising NHS services. Pledge Involvem
... -> We may detain you if we decide that you are suitable for detention and your application	can be considered	... whilst you are detained. You will be asked if there are any reasons why you should not
... epresent you in court, or you may hire your own. In some coun-tries, lawyers	may not be involved	... in the court hearing—the Central Authority will tell you if this is the case. Consular offic
... larly acquisitive of possessions. lack of The emotionally abused child: experiences	may include being scapegoated	... , deprived of physical affection, belittled and humiliated, repeatedly moved from significant relati

Les contextes étendus des structures [modal + passif canonique] dans les deux Figures révèlent tout d'abord qu'elles apparaissent souvent après des propositions conditionnelles en IF, WHERE ou WHEN/UNTIL, par exemple *If you are vulnerable and you are in Scotland, you may be allowed to have your screening interview in Glasgow* (8.2). En cela, ces contextes sont partiellement similaires aux textes LEX, qui utilisent également MAY et CAN pour exprimer les différentes possibilités

d'application et d'interprétation de la loi, rappelant ainsi une caractéristique paradoxale de la loi, qui vise à être universelle mais dépend également des cas individuels. Cependant, dans ces sous-corpus destinés au grand public, nous considérons que ces modaux, en particulier *may* expriment également une forme de *hedging* de la part des rédacteur.ices, pour indiquer aux usagers que le résultat indiqué dans la proposition ne sera peut-être pas réalisé. Ainsi, pour reprendre notre exemple, s'il est possible qu'une personne vulnérable en Écosse puisse être autorisée à obtenir un entretien à Glasgow, ce n'est pas toujours le cas, d'où la modalisation avec *may*. Les passifs qui entrent en collocation avec *will* contiennent également des précautions oratoires correspondant au *hedging*, avec des adverbes modalisants, comme *not necessarily* et *usually* : *it does not necessarily mean that you will be suspected of fraud, you will not usually be expected to remortgage* (Figure 8.1). Les rédacteur.ices ne peuvent en effet pas garantir que les contenus de la proposition principale, dans les conditions exprimées par la subordonnée conditionnelle en IF, s'appliqueront au justiciable dans sa situation personnelle. Pour Turnbull (2018), l'utilisation de la modalité et du *hedging* peut également venir renforcer la dimension communicative et présenter l'information avec politesse, en particulier car les conseils donnés par des expert.es pourraient être perçus comme menaçants (*face-threatening*) dans certains contextes anglophones. À l'inverse, le modal *can* ne semble pas avoir une fonction de *hedging*. On voit dans la Figure 8.2 qu'il est utilisé pour exprimer les différentes possibilités offertes par la loi (*This service*), et peut même avoir une valeur déontique, indiquant ce qui est interdit (ex : *Abducted children cannot be extradited*).

À travers différents degrés de modalité, les rédacteur.ices expriment donc une forme de positionnement par rapport au contenu présenté et le passif, associé à la modalité, semble donc avoir une fonction interpersonnelle. Pour le français, nous prenons l'exemple de FR-DILA pour étudier le passifs canoniques avec *devoir* et *pouvoir* à la troisième personne du singulier, en nous appuyant sur des extraits présentés dans la Figure 8.3.

Pour FR-DILA, les verbes *devoir* et *pouvoir* associés au passif ne semblent pas avoir une valeur de *hedging*. *Devoir* a une valeur déontique lorsqu'il est utilisé dans des constructions qui visent à expliquer les démarches judiciaires ou administratives à suivre dans certaines situations (un *signalement aux autorités* dans le cas d'abus de confiance et de maltraitance d'une personne âgée, par exemple), ou pour préciser le déroulement d'une démarche (*la plainte doit être déposée dans un délai de 6 mois*). Le passif permet de thématiser ces éléments légaux ou administratifs et de mettre en position focale l'information nouvelle qui détaille cette démarche. *Pouvoir*

FIGURE 8.3 – Passifs canoniques avec *pouvoir* et *devoir* dans FR-DILA

contre le curateur ou le tuteur. </s><s> La plainte doit	être déposée	dans un délai de 6 ans après la découverte des faits. </s>
elle fenêtre La réception de la plainte ne peut pas vous	être refusée	. </s><s> La plainte est ensuite transmise au procureur de
><s> À savoir : un signalement aux autorités doit aussi	être fait	en cas de maltraitance sur une personne âgée. </s><s> F
ré ou non). </s><s> Aucune pièce d'identité ne peut lui	être demandée	et aucune enquête ne peut être menée. </s><s> Elle peut
ne peut lui être demandée et aucune enquête ne peut	être menée	. </s><s> Elle peut lever le secret de son identité à tout m
, et des délais et conditions sous lesquels l'enfant peut	être repris	par ses parents. </s><s> Placement de l'enfant Lorsque l'
celui-ci est admis comme pupille de l'État et peut alors	être proposé	à l'adoption. </s><s> En revanche, si la mère revient sur é
médical). </s><s> Parents concernés Votre enfant peut	être inscrit	en halte-garderie quelle que soit votre situation familiale e
nt une rémunération. </s><s> Un contrat de travail doit	être conclu	entre vous (parent employeur) et l'assistante maternelle. <
e pas 5 heures consécutives. </s><s> Cet effectif peut	être réduit	lorsque l'accueil périscolaire est organisé dans le cadre d'
ocial ou des revenus. </s><s> À noter : le paiement des	sommes demandées	peut se faire au moyen de chèques emploi universel (Ces

est quant à lui employé pour exprimer les possibilités offertes par le droit selon les profils des justiciables, par exemple *vosre enfant peut être inscrit à la halte-garderie*, où *peut* est à entendre dans le sens de *a le droit*. Il semble donc que la modalité soit moins liée à des précautions oratoires dans ces extraits, peut-être parce que ce corpus provient d'un acteur institutionnel directement issu du gouvernement, dont on peut s'attendre à ce qu'il revendique un positionnement plus autoritaire. Cette différence pourrait également s'expliquer par une différence culturelle, les recommandations étant peut-être considérées comme moins menaçantes (*face-threatening*) en français, mais une analyse plus approfondie serait nécessaire pour vérifier cette hypothèse, peu d'études ayant appliqué la théorie des faces au contexte de la communication juridique⁵. Ces analyses suggèrent des différences de positionnement et d'engagement vis-à-vis du lectorat : l'ethos discursif construit dans FR-DILA ne semble pas atténuer la relation de pouvoir entre "sachants" et usagers non-experts.

La dimension interpersonnelle est en effet présente à différents degrés selon les sous-corpus, en particulier anglophones ; dans certains textes, elle peut dépasser la reformulation des connaissances juridiques. EN-Brochures contient de nombreuses formes passives de type <You (auxiliaire de modalité) + *be ask*>, et certains contextes révèlent que les rédacteur.ices vont plus loin dans l'expression de la solidarité avec le lectorat. Le premier extrait étudié (225) est destiné aux victimes de délits ou crimes sexuels ; l'extrait (226) est issu d'une brochure s'adressant à des demandeur.euses d'asile.

5. Voir cependant Kerbrat-Orecchioni (1991), Preite et al. (2007), qui étudient cette question dans les arrêts de la Cours de Justice de l'Union européenne, ainsi que Oursel (2012), qui analyse la gestion des faces dans les interactions entre agent.es de l'administration et usagers.

- (225) There are some court rules (law) about what they can and cannot ask you, and it is up to the judge to oversee these. **You cannot be asked** about your sexual experience with any person other than the defendant, except with the permission of the judge. **You cannot be asked** any question about your sexual reputation. There is no right or wrong way for you to handle cross-examination because each case is different. Some suggestions are : Listen carefully to the question, sometimes there can appear to be more than one question in what is being asked. Make sure you understand what you are answering. (. . .) If you don't understand a question, **ask for it to be explained**. If you don't hear a question, **ask for it to be repeated**. (EN-Brochures)
- (226) Your interview will take place in a Home Office building near to where you live. (. . .) When you arrive you will need to go through security. **This is nothing to be alarmed about** and is purely for safety reasons. **You will be asked** to remove any coats, jackets or belts and place them in a tray with the contents of your pockets. (EN-Brochures)

Dans ces extraits, le passif est utilisé pour thématiser les justiciable, *you*, à qui on explique pas à pas les étapes de la procédure qui les occupe. Dans le premier extrait, les rédacteur.ices explicitent d'abord les règles d'interrogatoire et de contre-interrogatoire dans le cas d'un procès pour viol ou agression sexuelle, anticipant peut-être des inquiétudes du lectorat quand au type de questions posées. Puis les rédacteur.ices prévoient et répondent à de potentielles difficultés à comprendre les questions, et encouragent donc à plusieurs reprises les lecteur.ices à faire répéter ou expliciter les questions en cas d'incompréhension, afin d'insister sur ce droit (*ask for it to be explained, ask for it to be repeated*). Dans l'extrait (226), qui s'adresse à des personnes demandant l'asile, c'est également les lecteur.ices qui sont placés en position thématique par *you*. On trouve également une phrase intéressante : *This is nothing to be alarmed about*. Cette proposition au passif est d'ailleurs assez complexe syntaxiquement, comme d'ailleurs les phrases de l'exemple (225) préconisant de faire répéter les questions lors d'un procès, puisque que toutes contiennent des propositions nominales en TO. Ces propositions présentent une focalisation de l'élément précédent, avec par exemple une reformulation générique en THIS en position thématique, suivie de l'information nouvelle, qui exprime une évaluation rassurante de la situation (*nothing to be alarmed about*). Or, ces propositions complexes sont tout de même utilisées pour anticiper une réaction affective, d'inquiétude ou de peur du justiciable, dans une situation qui comprend des aspects non seulement matériels, mais sont également très chargées émotionnellement.

Dans d'autres textes, extraits des corpus EN- et FR-Brochures, cette dimension interpersonnelle passe non par la deuxième personne mais par l'emploi des pronoms de première personne, visant une plus grande identification des lecteur.ices en anti-

cipant les questions qu'ils pourraient se poser (dont nous avons déjà parlé) :

- (227) Comment prouver que **je** suis une victime? Par un certificat médical **établi** par un médecin. Un médecin traitant ou les urgences de l'hôpital **sont habilités** à m'examiner.
- (228) What happens if **I am awarded** money or property at the end of the case? If you keep or gain money or property at the end of your case, you may have to repay some or all of your legal aid costs.

Loin d'être évité, le passif canonique participe au mouvement de personnalisation du texte. Lors du passage de LEX à PLAIN, on passe de sujets abstraits ou génériques, qui réfèrent principalement à des entités collectives ou à des principes juridiques (parfois des humains génériques), à des formes de deuxième personne s'adressent directement au lectorat non spécialisé, afin de recontextualiser la connaissance juridique dans leur propre expérience du monde. Le passif sert donc non seulement la dimension cognitive de la reformulation du contenu juridique (Turnbull 2018), mais peut également, à des degrés divers, une fonction interpersonnelle.

8.3.2 Dans les résumés de jugements, le passif pour thématiser les participant.es de la procédure judiciaire

Le sous-corpus EN-CA-Résumés contient également des constructions passives canoniques, bien qu'il soit le sous-corpus anglophone qui en contient le moins parmi ceux que nous proposons d'étudier. Il est de fait le sous-corpus le plus proche de certaines recommandations pour rédiger en PL, qui préconisent que 90% des formes verbales finies soient à l'actif (cependant ce chiffre ne prend pas en compte les passifs non-canoniques). Les passifs canoniques sont cependant bien présents dans ces textes, et semblent être utilisés dans des collocations spécifiques à la procédure judiciaire.

Les passifs canoniques sont fréquemment utilisés dans la partie du résumé qui rappelle les faits, en particulier pour thématiser les acteurs humains jugés (prévenus ou accusés), les affaires elles-mêmes ou les éléments de la procédure (*evidence*). Par exemple, on retrouve la collocation avec *find*, de la forme < [GN humain] was/were + found + (not) guilty >, que nous avons également vue dans EN-CA-Jugements.

- (229) Mr. Gubbins and Mr. Vallentgoed **were charged** with having blood alcohol "over 80" in separate incidents.
- (230) Without it, [Mr. Reeves **was found** not guilty].
- (231) Before [anyone **can be found** guilty] of a crime, a judge or jury has to believe that the person is guilty beyond a reasonable doubt.
- (232) The judge who heard the pre-trial argument agreed with Mr. Reeves that the computer evidence **couldn't be used**.

En français, le passif canonique est également utilisé pour rappeler les faits et thématiser les prévenu.es, soit comme individus nommés représentés par un nom propre, soit comme individus désignés par un nom à valeur générique (*une personne*); les principes juridiques sont également souvent sujets du passif.

- (233) M. Carson **a été accusé** de fraude envers le gouvernement pour s'être livré au trafic de son influence auprès de celui-ci.
- (234) Lorsqu'une personne **est arrêtée**, elle doit être informée des motifs de son arrestation et avoir la possibilité d'embaucher un avocat, ce qui n'a pas été le cas pour M. Penunsi.
- (235) La Cour suprême a jugé que la personne reconnue coupable a droit soit à la peine qui s'appliquait quand elle a commis son crime, soit à celle qui s'appliquait quand elle a **été condamnée** .

Le passif apparaît souvent employé avec *allow* et *autoriser*, afin d'explicitier les règles de droit, c'est-à-dire ce que les acteurs et actrices de la justice ont le droit de faire, par exemple comme on le voit ci-dessous des acteurs institutionnels collectifs, la police ou les jurés.

- (236) Judges can tell juries what kinds of inferences they **are allowed** to make.
- (237) Justice Abella said that, like other accused, police officers **should not be allowed** to share informer-privileged information with their lawyers unless they show they might be wrongfully convicted if they don't.
- (238) La Couronne (le poursuivant) a fait valoir que le policier **était autorisé** à prendre l'ordinateur partagé, car il avait eu la permission de la conjointe de M. Reeves.

Ces structures passives apparaissent souvent avec des auxiliaires de modalité ou des valeurs déontiques (*était autorisé*). En anglais en particulier, les collocats des formes de passif canonique sont *should* et *could*, comme le montre la Figure 8.4.

Dans ce sous-corpus, les modaux n'ont pas de valeur de *hedging* comme dans les textes plus administratifs. La dimension interpersonnelle est pourtant visible en

FIGURE 8.4 – Passifs canoniques et auxiliaires de modalité dans EN-CA-Résumés

he wrong rules to Ms. Awashish's request. </s><s> But this was a legal error, which could	be corrected	on appeal, not a jurisdictional error that needed to be fixed with certiorari. </s><s> So, eve
erty's basic rights would be permanently harmed. </s><s> It said certiorari should n't have	been granted	because the provincial court judge had the power to make the decision (even if she made
as told he would have to attend an immigration hearing when he was released, and could	be deported	. </s><s> Immigration rules also said he could not appeal the deportation for any reason. </
rt said that if Ms. Awashish asked for disclosure again, the documents' relevance should	be looked	at in light of the Court's decision in R. v. Gubbins, which was released on the same day. </
atabases were to the lawsuit, rather than on the kind of information the Act said could not	be shared	. </s><s> He also pointed out that information about "particular" insured persons did not m
ep the information that would identify customers anyway. </s><s> It said Rogers could n't	be compensated	for steps it took that overlapped with things it already had to do under the Act. </s><s> Jus
was unlawfully confined by Ms. Magoon and Mr. Jordan. </s><s> Therefore, the two could	be found	guilty of first-degree murder. </s><s> This case dealt with what constitutes unlawful confin
ed that the matter should go back to the motion judge to decide how much Rogers should	be paid	. </s><s> Police officers can't reveal information about confidential informers to anyone, in
watch people in nearby hotel rooms. </s><s> The security guards didn't think they should	be fired	. </s><s> Their union filed grievances (a formal complaint under a collective agreement) or
:</s><s> This settled a disagreement among lower courts about how these records should	be treated	. </s><s> R. v. Awashish, released on the same day, also dealt with breathalyzer records e
it thought financial markets were so important to the Canadian economy that they should	be overseen	by a single regulator. </s><s> In 2011, it asked the Supreme Court whether Parliament wo

anglais avec l'emploi de formes contractées et d'expressions courantes ou familières, toutes deux illustrées dans la phrase *The security guards didn't think they should be fired*. Le modal *could* est associé au passif pour rappeler ce qu'il était possible (ou non) de faire d'après la loi au moment de la procédure judiciaire ou du jugement de première instance, par exemple *the kind of information the Act said could not be shared*. Si c'est *COULD* qui est employé et non *CAN*, c'est que les structures passives se retrouvent dans des propositions nominales contenant du discours indirect (*reporting clauses*), rapportant soit le contenu de la loi, soit ce que les juges ou la Cour ont déclaré ou décidé. *Should* est également fréquemment employé pour rapporter ce qu'un juge ou une décision indique par rapport aux jugements de cours inférieures, par exemple *certiorari shouldn't have been granted*, qui rapporte un commentaire vis-à-vis d'une décision des cours inférieures. Les résumés imitent ainsi en partie les juges dont ils rapportent les paroles, puisque ceux-ci utilisent la modalité pour exprimer ce qu'ils pensent que la loi devrait être (Maley 1994). Le passif et les modaux rappellent donc l'hétérogénéité énonciative de ces textes, où les rédacteur.ices reformulent la voix des juges, qui eux-mêmes reformulent, en les interprétant, à la fois la loi et les acteurs des cours inférieures.

On note par ailleurs que dans tous les exemples cités, l'agent n'est pas exprimé : il s'agit donc de passifs courts sans compléments, qui font écho aux résultats de la littérature concernant l'absence de complément en *by* dans la plupart des passifs. Comme vu plus haut dans les textes administratifs, ce n'est que lorsque la question de l'agentivité est au cœur de l'affaire ou pourrait être instanciée par plusieurs acteurs selon les cas, que l'agent est précisé :

- (239) Les militaires accusés de crimes civils n'ont pas le droit d' **être jugés par un jury** s'ils **sont jugés par le système de justice militaire**, décide la Cour suprême.

Dans les résumés de jugement, d'après l'analyse des concordances, le passif canonique est utilisé pour thématiser les acteurs de la justice, en particulier prévenu.es et accusé.es, ainsi que les institutions collectives et leurs représentant.es (les juges ici). Dans ces exemples, à nouveau on voit que les rédacteur.ices effectuent un double mouvement de personnalisation et de généralisation. Cette fois, les rédacteur.ices n'utilisent pas de pronoms de deuxième personne, car ils ne s'adressent pas à leurs lecteur.ices, mais ils énoncent d'abord une généralisation, c'est-à-dire la décision prise par la CSC concernant la question de droit posée par l'affaire (comme on le voit dans l'exemple (233), qui constitue la première phrase du texte), avant de rappeler les faits spécifiques et individuels du cas et d'exposer le raisonnement des juges. Ces résumés reproduisent et explicitent ce que fait la Cour Suprême elle-même, en énonçant une règle de droit générale à partir d'une affaire particulière.

8.3.3 Passifs et fonctions textuelles dans le corpus PLAIN

Jusqu'ici, nous nous sommes intéressés aux passifs canoniques ayant une fonction référentielle. Or, on trouve des formes de passifs qui ont davantage une fonction métatextuelle.

Un des verbes les plus fréquemment passivés dans le sous-corpus EN-CAB est le verbe *call*. Comme nous l'avons vu dans le chapitre portant sur la terminologie, cette séquence semi-figée est souvent utilisée dans EN-CAB (351,2 pmm) pour expliciter des termes spécialisés dans une collocation de type < This is/These are/is + called + GN (terme) >. On trouve quelques occurrences de cette séquence dans EN-CA-Résumés (113,5 pmm), et son équivalent en français dans FR-FILA, <terme + est appelé(e)> :

- (240) If you have rent arrears, your landlord may try and evict you. This is called seeking possession.
- (241) When a large group of people have the same legal problem, they might decide to get together and sue as a group. **This is called** a class action. (En-CA-Résumés)
- (242) Le juge rend sa décision seul au vu du dossier présenté par le procureur de la République. Cette décision **est appelée** ordonnance pénale. (FR-DILA)

Ces structures suggèrent que le passif fait véritablement partie de la phraséologie de la médiation juridique, puisqu'elles semblent appartenir à des schémas lexico-grammaticaux qui ont pour fonction de définir et d'introduire des termes.

Dans le corpus EN-CA-Résumés, une collocation fréquente qui n'est pas liée à la procédure judiciaire, mais a davantage une valeur méta-linguistique, contient le verbe *mean*.

- (243) Long-term supervision **is meant to** protect the public. It **is also meant to** help long-term offenders re-enter society.
- (244) After juries hear all the evidence, they get instructions from the judge about how to decide if a person is guilty. **These are meant** to make sure juries make decisions based on the law and the evidence (not feelings or hunches).

La collocation <[Principe juridique] + *be meant* + to> sert à expliquer la fonction de certains éléments de la loi ou de certaines peines prononcées. On peut y voir une expression qui vient explicitement signaler au lectorat non-spécialisé l'intention qui se trouve derrière la législation ou derrière certaines règles.

Dans FR-DILA, tout comme la question "*De quoi s'agit-il ?*", mentionnée au Chapitre 7, permettait d'introduire la définition des termes, la question contenant un passif canonique "*Qui est concerné ?*" (187,1 pmm) a une fonction textuelle : elle signale aux justiciables que cette partie du texte les informe, en leur indiquant si le principe juridique dont il est question s'applique à leur situation.

- (245) **Qui est concerné ? Sont concernées** les majeurs : ayant fait l'objet d'une mesure d'accompagnement social personnalisé qui a échoué à rétablir leur autonomie dans la gestion de leurs ressources, et dont la santé et la sécurité sont de ce fait **menacées**, qui, par ailleurs, ne font pas l'objet d'une mesure de curatelle ou tutelle, et pour qui toute action moins contraignante (par exemple : application des règles relatives aux droits et devoirs du conjoint) s'avère insuffisante (...)

La question dans (245) est suivie d'un passif avec inversion sujet-verbe, correspondant à la même syntaxe que les textes législatifs cités plus haut. On retrouve une question directe au passif ayant la même fonction d'organisation textuelle, mais qui sert à introduire les termes dont il est question, dans EN-CAB (56.25 per million tokens), sous la forme <*What's meant by* + terme> :

- (246) *What's meant by inhuman treatment?* Inhuman treatment is ill-treatment which causes you severe mental or physical suffering.

Ces questions sous forme de propositions contenant des passifs canoniques ont une fonction rhétorique précise : ce sont des *discourse organizers*, qui allient à la fois une dimension cognitive d'explicitation de concepts du droit, et une dimension communicative puisqu'ils indiquent une anticipation des questions des justiciables.

Le passif semble donc remplir non seulement une fonction de ré-orientation du discours envers le lectorat, mais également une fonction textuelle, puisqu'on le retrouve en particulier dans des questions pour structurer le texte, ou des périphrases permettant d'explicitier les termes ou principes juridiques. Leur présence dans des séquences semi-figées récurrentes suggère l'utilisation d'une véritable phraséologie de la médiation juridique dans certains sous-corpus.

8.3.4 Passifs non-canoniques, cumul de formes passives et voix discursives dans le corpus PLAIN

Enfin, de nombreux exemples de passifs non-canoniques sont également présents dans le corpus PLAIN, notamment pour expliciter les termes :

- (247) Your appeal will be heard by an independent tribunal **called** the First-Tier tribunal. (EN-Brochures)
- (248) Anyone **detained** (held) in prison or at the Mangere Refugee Resettlement Centre or any other place **is said to** be in detention. (EN-Brochures)

Dans ces exemples, les passifs sans auxiliaire introduisent des termes ou acteurs institutionnels. Comme dans le corpus LEX, on trouve parfois des phrases contenant des "cascades" de passifs, c'est-à-dire des propositions, phrases et paragraphes contenant à la fois des passifs canoniques et sans auxiliaire à la suite.

- (249) À savoir : les actes d'état civil doivent être obligatoirement accompagnés de leur traduction en français par un traducteur interprète **agréé**.
- (250) L'affaire portait sur un concept juridique **appelé** "privilege". Celui-ci empêche certains types de renseignements **d'être révélés**. Le "secret professionnel de l'avocat" **est bien connu**, et protège les communications entre l'avocat et son client. (FR-CA-Résumés)
- (251) Anyone who can prove **they are owed** money can make a claim, **called** a "claim provable in bankruptcy." Anyone with a provable claim will get paid in a certain order set out in the BIA. (EN-CA-Résumés)

- (252) You **will be asked** a series of questions **tailored** to your individual circumstances to try and find out this information. (EN-CAB)

Comme pour le passif canonique, les *bare passives* apparaissent dans des propositions qui thématisent des référents humains :

- (253) A person **named** as an irrevocable beneficiary of a life insurance policy doesn't always have a right to keep the insurance money, the Supreme Court has confirmed.
- (254) Your landlord can ask the court for a warrant of possession ; you would then get a notice of eviction **sent** to you by the bailiffs saying when the eviction will take place. (EN-CAB)

Dans ces exemples, les passifs sans auxiliaire post-modifient des noms qui réfèrent soit à des concepts juridiques abstraits, soit à des documents juridiques (254), ou encore aux justiciables eux-mêmes (253). Comme pour le corpus LEX, il semble que les formes de passifs servent une véritable stratégie de présentation de l'information. En particulier, il semble utilisé pour introduire des termes à l'intérieur de certaines phrases, comme on le voit avec les exemples "un concept juridique **appelé** "privilège" ou *called a "claim provable in bankruptcy."* Il serait donc un équivalent plus concis des schémas LG *This is called* ou autres segments métatextuels qui introduisent les termes juridiques. Dans les autres exemples, les passifs sans auxiliaire viennent simplement post-modifier des noms et servent à la cohésion discursive au sein d'une même phrase, en permettant d'ajouter des précisions : ainsi les rédacteur.ices viennent-ils définir de qui il est question lorsqu'ils expliquent une décision (*anyone named as an irrevocable beneficiary*) ; ou encore à préciser les acteurs et l'origine de certains documents administratifs ou judiciaires (*a notice of eviction sent to you by the bailiffs*).

Concernant les autres formes de passif, on trouve dans le corpus EN-PLAIN quelques exemples de passifs avec un autre auxiliaire que *be, get*, dans chacun des genres discursifs représentés.

- (255) Discuss openly the effects and the risks of drugs, the illegal nature of the drugs and what it would mean if your child **got caught**.(EN-Brochures)
- (256) If you do this, you could **get fined** or even **sent** to prison. (EN-CAB)
- (257) Most cases **get appealed** to a court of appeal, but courts of appeal don't have the power to look at bail review decisions. (EN-CA-Résumés)

Cette utilisation des *get-passives*, réservée en anglais à un registre familier, souligne à nouveau l'oralité des textes PLAIN anglophone, avec une dimension interpersonnelle qui paraît plus appuyée. En français, l'utilisation d'un autre auxiliaire est exceptionnelle : on trouve un seul exemple, avec "*se trouver*" dans le corpus FR-PLAIN (*Un jeune de moins de 25 ans aura d'autres difficultés s'il se trouve isolé car il n'aura peut-être aucun revenu. (FR-Brochures)*). On trouve cependant dans FR-PLAIN diverses formes de passifs non-canoniques.

Dans FR-CA-Résumés, le sous-corpus qui contient le moins de formes canoniques, les rédacteur.ices ne sont pas avares de passifs sans auxiliaire ou de passifs non-canoniques, comme le montrent ces exemples :

- (258) L'alinéa 11i) indique ce qui arrive quand la peine **prév**ue pour un crime **est modifiée** entre le moment de la perpétration du crime et celui de la sentence **prononcée** pour ce crime.
- (259) Il **n'est pas permis** de présenter au procès des éléments de preuve susceptibles d'appuyer ce mythe, sauf s'ils sont pertinents et utiles à d'autres fins importantes.

L'exemple (258), qui contient un passif canonique, est une phrase de 32 mots, bien au-dessus de la longueur des phrases recommandée pour un style clair, rallongée notamment par deux *bare passives* et un nom prédicatif (*perpétration*). On peut d'ailleurs se demander pourquoi les rédacteur.ices n'ont pas préféré un passif canonique : (...) *entre le moment où le crime est commis et celui où la sentence est prononcée*. Dans l'exemple (259), l'emploi du passif impersonnel en *il* est contraint, comme dans EN-CA-Jugements plus haut, pour des raisons syntaxiques : la proposition nominale *présenter au procès des éléments de preuve susceptibles d'appuyer ce mythe* forme un sujet trop long pour être placée en position thématique, qui entraînerait une syntaxe trop lourde (*top heavy*). Ce sous-corpus, qui semblait à première vue suivre de près les recommandations pour un style clair, présente donc en réalité une forme de complexité syntaxique dans les GN qui sont post-modifiés par propositions passives réduites (sans auxiliaire). C'est également le cas dans FR-DILA, où les passifs sans auxiliaire apparaissent également fréquemment dans les textes. Certains passages, de par l'utilisation de ces *bare passives* associés à des noms prédicatifs, semblent rester très proches du style juridique, comme l'extrait (260) :

- (260) La réparation est équivalente au montant du prix du bien **détourné**, auquel peuvent s'ajouter une indemnité **destinée** à couvrir : le montant des frais **engagés** pour le procès, les frais **occasionnés** par la privation de l'objet, le préjudice moral. (FR-DILA)

Cette phrase, issue du corpus FR-DILA, comprend non seulement des *bare passives* mais de leur association avec des mots verbes relativement rares (*occasionnés, engagés*), qui rappellent les résultats du chapitre 7. La voix discursive, c'est-à-dire l'image rhétorique construite par les rédacteur.ices, n'est donc pas caractérisée par une forme d'informalité et d'oralité comme le suggère l'utilisation des passifs en *get* en anglais.

Que suggèrent ces exemples, globalement, sur le passif dans ces corpus destiné aux non-spécialistes ? Comme dans le corpus LEX, le cumul de constructions passives diverses dans le corpus PLAIN semble relever d'une stratégie générale de présentation de l'information. Les passifs sans auxiliaire sont souvent inclus, comme différentes types de propositions enchâssées (relatives, complétives), pour ajouter une précision, expliciter un terme ou encore introduire un circonstant pour préciser une situation liée à la procédure en question.

8.4 Résumé du Chapitre 8

Dans ce chapitre, nous nous sommes efforcées de caractériser l'utilisation du passif dans les corpus LEX et PLAIN. Si les résultats quantitatifs concernant les passifs canoniques suggéraient des différences entre les registres et les genres discursifs, l'étude qualitative révèle que les corpus utilisent tous des formes canoniques et non-canoniques de passifs, en particulier des *bare passives* et, pour le français, des passifs impersonnels.

Dans le corpus LEX, le passif, sert à thématiser principalement les acteurs institutionnels et principes abstraits du droit. Sous toutes ses formes, il paraît, en français comme en anglais, intrinsèquement lié au style nominal favorisé par les rédacteur.ices de textes législatifs et judiciaires. Ainsi semble confirmée l'idée de Minton (2015) pour ce qui concerne les textes médicaux : selon l'auteur, le rôle du passif n'est pas d'occulter l'Agent (dont on a vu qu'il est le plus souvent évident), mais de maintenir la configuration stylistique (*stylistic patterns*) dans la présentation de l'information. Or, notre étude des schémas LG semble bien montrer que le passif est une des formes caractéristiques de la phraséologie du droit.

Concernant le corpus PLAIN, le résultat principal est que tous les textes, y compris les sous-corpus anglophones qui se réclament explicitement du *plain*

language, utilisent des formes de passif canoniques et non-canoniques. Cette analyse suggère que le passif, bien que critiqué et déconseillé dans les recommandations, remplit une véritable fonction discursive, et qu'il participe à la ré-orientation du contenu juridique. Associé à d'autres stratégies typiques de la médiation juridique (comme l'emploi de questions directes ou de la deuxième personne), il sert à recontextualiser l'information juridique en thématissant les justiciables et les acteurs humains du droit. Dans sa forme sans auxiliaire, il est utilisé pour préciser les concepts et acteurs dont il est question en ajoutant cependant une forme de complexité syntaxique. Ce résultat est cohérent avec les reformulations en style clair étudiée par exemple dans le domaine médical (Gledhill et al. 2019), et avec les résultats des chapitres précédents. L'étude des contextes a également confirmé les différences de positionnement et d'engagement (*engagement*) des rédacteur.ices vis-à-vis du lectorat entre les sous-corpus, notamment entre le français et l'anglais dans les textes administratifs.

Chapitre 9

Distribution des procès sémantiques dans un échantillon de corpus portant sur la discrimination

Nous poursuivons notre étude des stratégies discursives de la diffusion du droit en nous penchant sur des échantillons plus réduits de nos corpus LEX et PLAIN. A partir d'extrait de corpus, nous proposons une étude de cas des procès sémantiques dans des textes juridiques et leurs versions simplifiées, et plus généralement la notion de transitivité telle qu'elle est conçue dans le modèle GSF. Dans ce chapitre, nous présentons la distribution des procès sémantiques des textes avant et après simplification, que nous allions à une analyse qualitative. L'échantillon de corpus de textes législatifs est composé d'un extrait de l'*Equality Act* de 2010 pour l'anglais, et d'extraits du Code pénal pour le français (de France), mis au regard de leurs versions PLAIN, issues de EN-CAB et FR-DILA. Concernant le discours judiciaire, l'échantillon de corpus est constitué des motifs du jugement majoritaire de l'affaire Centrale des syndicats du Québec v. Québec (Attorney General), motifs rendus par la Juge Abella, en français et en anglais, pour une affaire qui porte sur la discrimination salariale envers les femmes (la décision rendue est le rejet du pourvoi); leur version simplifiée consiste en deux résumés dans chaque langue qui portent sur cette affaire (deux jugements ont été rendus ce jour-là).

Nous rappelons que dans le modèle de la GSF, les procès sémantiques appartiennent au système de la transitivité, et correspondent à la métafonction expérien-

tielle, qui est définie ainsi : blent e

" [The experiential metafunction is] the "content" function of language; it is the language as the expression of the processes and other phenomena of the external world, including the world of the speaker's own consciousness, the worlds of thought, feeling and so on" (Halliday 1978 : 48).

La présentation d'un procès dans une proposition est décrite selon plusieurs éléments principaux dans le système de la GSF (Banks 2011), désignés par des termes que nous employons dans la suite de ce chapitre : le procès lui-même, généralement (mais pas toujours) réalisé par le verbe, et correspondant à un type sémantique ; et les participants qui prennent part au procès (c'est-à-dire dans les termes, de la grammaire traditionnelle, les arguments du verbe) ; éventuellement, les circonstances du procès.

En particulier, nous nous intéressons aux types de procès sémantique, et donc à la manière dont la représentation de l'expérience du monde physique (externe) et psychique (interne) est réalisée par les ressources grammaticales. Pour coder les procès sémantiques, nous utilisons une version simplifiée des catégories de la GSF, comprenant quatre catégories principales de procès sémantiques : matériels, relationnels, mentaux et verbaux. Il faut reconnaître que la typologie que nous utilisons est assez restreinte, et qu'il existe des sous-types pour ces catégories. Par exemple, les procès verbaux sont un sous-type de procès mentaux, qui contiennent d'autres sous-types : *show*, *prove* sont par exemple des procès verbaux démonstratifs (liés à la démonstration), tandis que des verbes comme *experience* dénotent des procès perceptifs. De même, les procès relationnels peuvent être des procès d'attribution ou d'identification, et les procès matériels incluent notamment des procès comportementaux (Halliday et Matthiessen 2013). Cependant, nous avons choisi une catégorisation minimale afin de simplifier l'analyse et d'observer globalement si certains types de procès sont davantage représentés dans certains textes de nos échantillons de corpus.

Nous présentons d'abord la distribution des procès sémantiques dans les extraits de corpus, en comparant les genres discursifs appartenant au même registre entre eux, puis les textes spécialisés avec leurs versions simplifiées (Sections 9.1 à 9.3). Dans un second temps, nous nous concentrons plus particulièrement sur les prédicats complexes et à la modalité dans ces échantillons de corpus. Il nous semble en effet intéressant d'examiner si ces constructions sont présentes dans la version en style clair, c'est-à-dire si le procès est exprimé par le prédicat ou par un nom, puisque nous avons vu que ces constructions sont caractéristiques des textes juridiques. La

présence d’auxiliaire de modalité est également un objet d’attention, puisqu’elle révèle différents positionnements des rédacteur.ices, comme nous l’avons vu.

Dans la suite de ce chapitre, lorsque nous présentons des exemples, nous indiquons entre crochets le prédicat [Prédicat] et le type de procès [Procès]. Nous précisons le type de procès à l’aide aux abréviations suivantes : [Procès : mat] pour les procès matériels ; [Procès : rel] pour les procès relationnels ; [Procès : men] pour les procès mentaux et [Procès : verb] pour les procès verbaux.

9.1 Résultats quantitatifs : distribution des procès sémantiques dans les échantillons de corpus

Dans cette section, nous présentons les résultats de la distribution des procès sémantiques en % du nombre de verbes codés dans les échantillons. Ces procès sémantiques ont été codés à la main, tout comme les prédicats complexes et les modaux. Les procès sémantiques sont présentés par couleur dans l’Annexe D. Notons que certains verbes peuvent être considérés comme réalisant plus d’un seul type de procès car ils sont sémantiquement ambigus ; c’est pourquoi certains verbes sont codés de deux couleurs différentes. Ils sont alors comptés deux fois.

9.1.1 Résultats pour l’anglais

Les résultats quantitatifs concernant la distribution des procès sémantiques dans les échantillons anglophones sont présentés dans le Tableau 9.1, et permettent déjà de dégager des différences entre les différents genres représentés.

TABLEAU 9.1 – Distribution des types de procès, des modaux et des prédicats complexes dans les textes anglophones (en % du nombre de verbes codés)

	Equality Act	Motifs Abella EN	EN-CA-Résumés	EN-CAB
Procès matériels	38	36,3	43,1	30
Procès relationnels	46,3	37	21,3	48,2
Procès mentaux	6,5	18,7	20,6	13,2
Procès verbaux	9,2	8,1	15	8,7
Prédicats complexes	13,6	3,7	1,3	6,1
Auxiliaire modaux	7,6	7,6	4,4	12,4
Nombre de procès codés	522	568	160	380

Nous comparons ici les échantillons de textes destinés au même lectorat entre

eux, c'est-à-dire l'*Equality Act* vs. les motifs de jugement, et les résumés vs. les textes d'EN-CAB. Dans les sections suivantes, nous comparons les textes spécialisés à leur version simplifiée.

Pour les textes spécialisés, le Tableau 9.1 suggère tout d'abord que les deux textes juridiques contiennent à peu près la même proportion de procès matériels et de procès verbaux. En revanche, on voit distinctement que les procès relationnels sont sur-représentés dans l'extrait de l'*Equality Act*, avec plus de 45% de procès relationnels, alors que la proportion est la même pour les procès relationnels que pour les procès matériels dans les motifs de jugement (environ 37%). Dans l'extrait de jugement, ce sont les procès mentaux qui sont bien plus fréquents : l'extrait de jugement contient trois fois plus de procès mentaux que le texte de loi. Quels sont les facteurs qui expliquent ces différences ? Les textes de lois ont pour fonction de définir les relations de droit entre les justiciables, et ici cette différence suggère que le texte législatif favorise les procès relationnels pour définir juridiquement la discrimination. D'une part, les procès relationnels d'identification sont utilisés dans des passages méta-textuels (dont nous avons déjà parlé) qui définissent les termes utilisés dans la suite du texte de loi :

- (261) The authorities to which this section **applies** [Procès : rel] are [Procès : rel] a Minister of the Crown ;
a government department other than the Security Service, the Secret Intelligence Service or the Government Communications Head quarters (...)

D'autre part, des procès relationnels d'attribution et d'identification précisent les caractéristiques protégées par la loi, comme dans l'exemple (262), la caractéristique qui correspond aux personnes mariées ou dans une union civile :

- (262) A person [Medium] has [Procès : rel] the protected characteristic of marriage and civil partnership [Attribut] if the person [Medium] is [Procès : rel] married [Attribut] or is [Procès : rel] a civil partner [Attribut].

La création de ces relations juridiques entre les justiciables et les caractéristiques (incluant l'âge, la race, le genre, etc.) semblent expliquer la grande proportion de procès relationnels dans cet extrait de l'*Equality Act*. Concernant la plus grande proportion de procès mentaux dans l'extrait de jugement, elle provient d'une différence fondamentale entre les genres législatifs et les jugements. Les textes législatifs

sont des textes performatifs et normatifs, essentiellement déclaratifs. Les décisions de la Cour Suprême, à l'inverse, s'ils sont partiellement narratifs et normatifs, sont surtout argumentatifs. Les juges, en rendant une décision motivée, expliquent leur raisonnement par rapport à celui des cours inférieures et des parties grâce à des procès mentaux. Ici en particulier, la Juge Abella explique comment les pouvoirs exécutifs de la province du Québec ont cherché une solution pour résoudre l'inégalité salariale entre hommes et femmes, et comment le juge de première instance a analysé les mesures prises par le Québec (263).

- (263) Quebec was [Prédicat] **seeking** [Procès : Men] a wide-ranging and pioneering approach to pay equity for women in workplaces without male comparators.
- (264) The trial judge **reviewed** [Procès : Men] in considerable detail the reasons for the delay and **found** [Procès : Men] that it was [Prédicat] not caused [Procès : Mat] by a lack of diligence on Quebec's part. (EN-CA-Jugements)

Dans l'extrait (263), on voit que la juge relate d'abord l'intention du Québec (agent collectif par métonymie renvoyant à des référents humains) avec le procès *seeking*. Par la suite (264), elle explique l'analyse qu'a faite le juge de première instance via deux procès mentaux, *review* et *find*, qui font référence à l'interprétation lors du premier procès, concernant le retard de l'application de la loi imposant l'égalité salariale entre hommes et femmes (représenté par un procès matériel *cause a delay*). Les procès mentaux sont donc prégnants dans les motifs de jugement, puisque ces textes expliquent à la fois le raisonnement des juges et celui des cours inférieures. Une autre raison qui explique peut-être la plus grande incidence des procès mentaux dans cet extrait est que la juge fait référence à la discrimination en tant qu'elle a un impact sur les personnes qui la subissent :

- (265) There is no doubt that the claimants in this case will **experience** [Procès : Mental] a considerable economic impact as a result of the delay.

Le texte contient en effet plusieurs occurrences de verbes exprimant des procès mentaux comme *suffer* et *experience* pour faire référence aux conséquences ressenties de la discrimination. Une autre divergence notable qui apparaît dans le Tableau 9.1 entre les deux genres juridiques est celle des prédicats complexes, qui sont presque trois fois plus fréquents dans le texte législatif que dans le jugements. Cela laisse supposer qu'ils sont exploités plus souvent dans le discours législatif que judiciaire

d'après ces échantillons, mais restent tout de même très élevés dans plusieurs types de discours "élaborés", puisque nous avons vu de nombreux exemples dans les corpus de jugements dans les Chapitres précédents. Nous en discutons de ces prédicats complexes plus loin en comparant ces constructions au français (Section 9.3).

Nous passons maintenant à la description des résultats quantitatifs dans les textes de médiation juridique anglophones. Le Tableau 9.1 révèle également des différences notables entre les deux genres discursifs. Les trois textes du sous-corpus EN-CAB contiennent presque 50% de procès relationnels (comme le texte législatif correspondant), pour seulement 22,5% dans le corpus EN-CA-Résumés. À l'inverse, ce dernier contient beaucoup plus deux fois plus de procès mentaux et environ un tiers de plus de procès verbaux que EN-CAB. À nouveau, ce résultat semble dû à une différence entre les deux genres discursifs. Les résumés rapportent les raisonnements et les paroles des parties et des juges grâce à des procès mentaux et surtout verbaux, comme illustré dans l'exemple (266) ci-dessous. Les textes sur la discrimination dans EN-CAB visent quant à eux à expliquer ce que constitue une discrimination au regard de la loi par le biais de définitions réalisées par de procès relationnels, comme on le voit dans l'exemple (267).

- (266) Chief Justice Beverley McLachlin (as she **was** when the case was **heard** [Procès : men]) **agreed** [Procès : verb] with Justice Abella that the women's equality rights had been breached. But she **disagreed** [Procès : verb] that this breach was constitutionally **justified**. (EN-CA-Résumés)
- (267) Direct discrimination **is** [Procès : rel] against the Equality Act 2010. Discrimination which **is** [Procès : rel] against the Equality Act **is** [Procès : rel] unlawful. This **means** [Procès : verb] you can take action in the civil courts. (EN-CAB)

Dans l'exemple (266), nous avons souligné les procès verbaux et mentaux, dont on voit qu'ils sont sur-représentés dans cet extrait, même s'ils sont parfois au passif : ils expriment des procès communicatifs (*agree, disagree*) ou des processus mentaux (*hear, justify*) liés aux délibérations et argumentations des juges. L'extrait d'EN-CAB s'efforce quant à lui de définir la discrimination directe pour relier la situation personnelle du justiciable à des situations de discrimination, grâce la répétition de procès relationnels attributifs. Les rédacteur.ices précisent aussi qu'une telle situation signifie que des recours judiciaires sont possibles grâce au procès verbal *mean*.

9.1.2 Résultats pour le français

Nous passons maintenant aux résultats obtenus pour le français, qui sont présentés dans le Tableau 9.2.

TABLEAU 9.2 – Distribution des types de procès, prédicats complexes et auxiliaires modaux dans les échantillons en français (en % du nombre total de verbe)

	FR-Lois	FR-Jugements	FR-Résumés	FR-DILA
Procès matériels	38	38	41,2	52,5
Procès relationnels	23,5	34,1	26,1	24,9
Procès mentaux	20,9	18,8	15	12,8
Procès verbaux	17,6	8,9	17,6	9,8
Prédicats complexes	5,9	7,3	5,9	9,1
Auxiliaires modaux	3,2	6,3	2	26,8
Nb procès codés	187	671	153	265

Comme pour l'anglais, nous comparons dans cette section les textes réputés complexes issus de FR-LEX, puis les échantillons de textes du corpus FR-PLAIN entre eux.

Les deux échantillons de corpus spécialisés contiennent tous deux une majorité de procès matériels, qui composent environ 40% des verbes codés. Les procès mentaux sont à peu près dans les mêmes proportions. Des différences apparaissent en revanche dans la distribution des procès relationnels, avec respectivement 23,5 % dans les textes de lois français contre et 34,1 % dans les motifs de jugement. Cette différence vient sans doute de la plus grande proportion de propositions qui posent la question de la discrimination dans les motifs de jugements par le biais de procès relationnels, par exemple lorsque la juge se demande "s'il y **avait** un distinction **fondée** sur le sexe (...)" (FR-CA-Jugements). Surtout, un des résultats qui peut paraître étonnant est qu'il y a une nette différence dans la distribution des procès verbaux, qui sont deux fois plus fréquents dans les articles de lois français. On aurait en effet pu s'attendre à ce que ces procès soient plus fréquents dans les motifs de jugement, comme nous l'avons vu pour l'anglais, puisque les juges de la CSC rapportent les propos des cours inférieures et des parties, mais ce n'est pas le cas dans nos échantillons. Or, dans les textes français présents dans l'échantillon, on trouve l'article de loi qui définit l'infraction et les peines liées au délit spécifiquement français d'incitation à la haine.

(268) Tous cris ou chants séditieux **proférés** [Procès : verb] dans les lieux ou réunions publics seront punis [Procès : mat] de l'amende prévue pour les contraventions de la 5ème classe. Ceux qui, par

l'un des moyens **énoncés** [Procès : verb] à l'article 23, auront **provoqué** [Procès : verb/mat] à la discrimination, à la haine ou à la violence à l'égard d'une personne ou d'un groupe de personnes (...), seront punis [Procès : mat] d'un an d'emprisonnement et de 45 000 euros d'amende ou de l'une de ces deux peines seulement.

Ce délit fait explicitement référence à des paroles prononcées par les personnes en infraction pour inciter à la haine ou à la discrimination, ce qu'on voit avec le verbe *proférer*. Nous pouvons nous arrêter un instant sur le verbe *provoquer* dans l'exemple (268). Il réalise dans ses emplois typiques un procès à la fois matériel, car il affecte bien un objet et a des conséquences matérielles (discrimination, actes de violences). Cependant, ce verbe est selon nous ambigu dans ce contexte et peut également être considéré comme un Procès verbal, puisque, dans le cadre du délit d'incitation à la haine ou à la discrimination, il est explicitement lié à des paroles ou écrits, comme le précise le texte de diffusion du droit : "L'incitation à la haine raciale concerne des **propos** haineux ou violents" (FR-DILA). C'est pour cette raison que ce verbe a été codé deux fois, comme procès mental et comme procès verbal, illustrant les chevauchements possibles entre les catégories des procès sémantiques. Ce délit particulier, spécifique au droit français et dont il n'est pas question dans les motifs de jugement (qui porte sur la discrimination salariale), explique donc en partie cette différence quantitative entre les deux échantillons. On remarque également dans l'exemple (261) la présence du verbe métatextuel *énoncer*, sous forme de passif périphrastique qui fait référence à un autre article du texte, que nous considérons également comme un procès verbal puisqu'ici on fait référence à ce que les législateur.ices ont dit (*énoncé*) dans un autre passage de la loi. On trouve dans les extraits de FR-Lois quatre occurrences de *énoncé(s)* et quatre de *mentionné(s)* pour renvoyer à des articles de la loi, ce qui explique également la plus grande proportion de procès verbaux.¹

En ce qui concerne la distribution des types de procès sémantiques dans les extraits du corpus FR-PLAIN, à l'inverse, c'est les textes de FR-CA-Résumés qui contiennent davantage de procès verbaux que les textes de FR-DILA. En voici quelques exemples :

- (269) Les juges Suzanne Côté, Russell Brown et Malcolm Rowe **ont exprimé** [Procès : verb] leur désaccord. Ils ont **affirmé** [Procès : verb] que la loi québécoise ne portait pas atteinte aux droits des femmes à l'égalité. Ils ont **souligné** [Procès : verb] que les droits conférés par la Charte sont

1. Nous nous concentrons dans cet extrait sur les procès verbaux, mais on note également les procès matériels au passif *seront punis* suivi de la peine.

"négatifs", en ce sens que l'État n'est pas tenu d'adopter des lois particulières — mais que, s'il en adopte, les lois en question doivent être conformes à la Charte. (FR-CA-Résumés)

Dans cet exemple, les rédacteur.ices du résumé expliquent les opinions dissidentes des juges qui ne se sont pas rallié à l'opinion majoritaire et ont rendu d'autres motifs. Pour cela, plusieurs verbes exprimant des procès verbaux sont utilisés avec ou sans projection d'une proposition nominale (*affirmé que* + proposition, *souligné que* + proposition vs. *exprimé leur désaccord* COD). On peut estimer que c'est parce que ces textes rapportent explicitement les paroles des juges qu'ils contiennent davantage de procès verbaux que les textes de FR-DILA, qui viennent surtout guider les justiciables en leur expliquant quoi faire en cas de discrimination. Cela explique sans doute pourquoi dans les textes issus de FR-DILA, les procès matériels sont sur-représentés avec plus de 50% des verbes codés appartenant à cette catégorie, ce sur quoi nous reviendrons. Dans les deux échantillons, on trouve cependant la même proportion de procès relationnels et mentaux. Nous allons maintenant nous intéresser plus en détail à ces différents types de procès en comparant les textes en style clair à leur version non-simplifiée.

9.2 Comparaison des procès sémantiques dans un jugement de la CSC et deux résumés

Après avoir comparé les genres du même registre entre eux, nous confrontons les textes juridiques avec les échantillons de textes simplifiés pour examiner illustrer qualitativement les éventuelles différences dans la distribution des procès sémantiques avant et après simplification.

9.2.1 Dans les motifs et résumés en anglais

Les échantillons de EN-CA-Jugement et EN-CA-Résumés étudiés contiennent à peu près la même proportion de procès mentaux d'après le Tableau 9.1, afin de rapporter les décisions et raisonnements des parties (en particulier le gouvernement du Québec qui devait déterminer l'écart salarial et trouver des solutions), mais également l'analyse des juges eux-mêmes. La proportion de procès matériels est également similaire dans les deux échantillons. Ces procès sont souvent utilisés au début des résumés de jugements, juste après l'introduction, pour raconter des faits

survenus dans le cadre de l'affaire, comme on le voit dans l'exemple, qui est la deuxième phrase du résumé.

- (270) To **tackle** [Procès : mat] the problem of wage discrimination against women workers, Quebec **passed** [Procès : mat] a law in 1996 **forcing** [Procès : mat] employers with 10 or more employees to **give** [Procès : mat] equal pay for work of equal value. (EN-CA-Résumés)
- (271) By **restricting** [Procès : mat] pay equity to every five years, the law **let** [Procès : mat] employers off the hook. (EN-CA-Résumés)

On voit que l'agent dans ce cas est soit un acteur institutionnel collectif - *Quebec* (270), *employers* (271) soit une entité abstraite comme la loi elle-même (271). Dans des affaires pénales, notamment relatant des faits de meurtres ou d'agression, les Agents et procès sont généralement plus concrets² mais ce n'est pas le cas ici, l'affaire portant sur une loi passée par un gouvernement fédéral. Dans les jugements, l'agent de ces procès matériels est souvent très abstrait et a la forme d'un GN contenant de nombreuses post-modifications. Le procès sémantique est quant à lui souvent réalisé par des groupes verbaux au passif, canonique ou sans auxiliaire (ce qui n'est pas sans rappeler les résultats du Chapitre 8), comme le montre l'exemple ci-dessous :

- (272) The two-year grace period **provided** [Procès : mat] by 38 further **postponed** [Procès : mat] pay equity for workplaces without male comparators until May 5, 2007. As a result, access to pay equity under the Act for women in workplaces without male comparators was **delayed** [Procès : mat], in accordance with 38, by two years beyond the length of time it took to **enact** the Regulation under 114. (EN-CA-Jugements)

Cet exemple illustre bien la dimension très abstraite des agents des procès matériels et le fait que les GN sujets qui les représentent sont très longs, comme *access to pay equity under the Act for women in workplaces without male comparators*. On note également que la deuxième phrase de l'extrait compte deux prédicats complexes. Le premier est le prédicat complexe lexicalisé *take time*, où le nom prédicatif exprime la portée du procès et réalise un procès matériel ; le second est *enact the Regulation*, composé du verbe et de la nominalisation de *regulation*, qui est également un procès matériel et forme cette fois une collocation spécialisée. Ces prédicats complexes sont par ailleurs insérés dans des structures syntaxiques complexes, puisque le premier

2. Par exemple dans la phrase suivante, tous les procès sont des procès matériels concrets : *Mr. Suter and his wife were arguing as he pulled up to park in front of the restaurant. At one point, he accidentally hit the gas pedal instead of the brake and drove into the patio.* (EN-CA-Résumés)

est dans une proposition relative, *time it took*, qui post-modifie un premier GN qui se trouve dans un GP, tandis que le second fait partie d'une proposition infinitive qui elle-même post-modifie le GN contenant la relative. Ces éléments syntaxiques et les prédicats complexes montrent bien à nouveau que le discours juridique favorise l'*intricacy*, c'est-à-dire l'interaction complexe entre plusieurs structures syntaxiques, ici des prédicats complexes à l'intérieur de plusieurs propositions hypotactiques. Par comparaison, les extraits (270) et (271) sont bien plus concis, les GN sujets étant également plus courts et plus concrets, bien qu'il y ait également une forme d'hypotaxe, limitée cependant à une ou deux propositions dépendantes par phrase. Les faits sont exposés sans entrer dans les détails (contrairement à l'extrait de jugement, qui cite des dates et articles de loi), suggérant une ré-organisation de la présentation de l'information dans les résumés, qui appartiennent à un genre discursif qui réclame, par définition, de la concision. On remarque également dans l'exemple (271) un procès matériel, *let someone off the hook*, réalisé par une expression idiomatique informelle, qui montre également un changement de niveau de langue dans le passage de EN-CA-Jugements à EN-CA-Résumés.

On trouve d'autres différences notables entre les deux échantillons de corpus. Par exemple, les procès relationnels dans les motifs de jugement rendus par la juge Abella sont plus fréquents que dans les deux résumés. Cette différence nous semble due à la nature argumentative des motifs de jugement, comme on le voit dans l'exemple ci-dessous.

- (273) The essence of Quebec's argument **is** [Procès : rel] that the pressing and substantial objective **was** [Procès : rel] to enact a scheme that created an effective remedy for systemic pay discrimination for women who were outside the existing pay equity scheme.
- (274) The next question **is** [Procès : rel] whether the delay **is** rationally connected to the purpose.
- (275) The test **is** [Procès : rel] whether the balance Quebec struck **falls within** [Procès : rel] the range of "reasonable alternatives" (para 160) and **results** [Procès : mat] in greater benefit than harm.

Dans ces phrases, la juge utilise des procès relationnels d'identification précisément pour identifier les questions de droit posées par l'affaire auxquelles elle va répondre, et la nature des arguments des parties. Les sujets de ces verbes réalisant des procès relationnels, comme *essence*, *objective* (273), *question* (274) ou *test* (275), sont des "noms sémiotiques" (*semiotic nouns*), c'est-à-dire des noms abstraits généraux qui expriment des notions mentales et communicationnelles (Benítez Castro et al. 2014), dont on peut considérer ici qu'ils ont une fonction cataphorique, en annonçant les étapes de la démonstration. Ces phrases servent donc à organiser le

discours, et les procès relationnels ont une fonction de cohésion et d'organisation textuelles : il permettent d'exposer pas à pas les étapes de l'argumentation de la Juge. L'exemple (274) montre par ailleurs qu'il peut y avoir des cas de verbes ambigus puisque le verbe *connected* n'est pas un prédicat dans ce contexte, mais l'attribut du verbe *be*.

Si le jugement contient plus de procès relationnels, à l'inverse, on trouve davantage de procès verbaux dans EN-CA-Résumés que EN-CA-Jugements. En effet, comme dans l'exemple en français évoqué plus haut, on observe dans les résumés de jugement anglophones une variété de verbes de paroles (*reporting verbs*), qui servent à introduire les arguments des juges et expliciter les différents points saillants de leurs motifs :

- (276) Justices Suzanne Côté, Russell Brown, and Malcolm Rowe **disagreed** [Procès : verb]. They **said** [Procès : verb] Quebec's law did not breach women's equality rights. They **noted** [Procès : verb] that Charter rights are "negative rights," **meaning** [Procès : verb] that governments do not have to adopt particular laws—but if they do, those laws must comply with the Charter.

Alors que dans les motifs, les juges expriment leur propre opinion, les résumés introduisent chacun des arguments des juges qu'ils introduisent par une proposition principale comportant un prédicat qui lui réalise un procès communicatif, introduisant une proposition complétive contenant le discours rapporté. A nouveau apparaît donc l'hétérogénéité énonciative explicite caractéristique de ce sous-corpus, déjà évoquée aux Chapitres 7 et 8. Ce résultat concernant les procès verbaux semble cohérent avec l'idée qu'on trouve dans ces résumés une ré-orientation du discours vers les référents humains, qui passe en particulier dans ce contexte par la figure centrale des juges en tant que figure d'autorité et énonciateur.ices du discours spécialisé de départ.

9.2.2 Dans les motifs de jugement et résumés en français

Les échantillons de jugements et les résumés anglophones et francophones étant co-rédigés et constituant donc la traduction l'un de l'autre, il est logique de trouver les mêmes différence et similarités dans la distribution des procès sémantiques. Nous les décrivons brièvement. Ainsi les motifs de jugement et les résumés contiennent-ils à peu près la même proportion de procès matériels (environ 40%), pour rappeler les faits et en particulier l'impact du retard de l'application de la loi sur les parties.

On trouve, comme en anglais, davantage de procès relationnels dans les jugements, mais deux fois plus de procès verbaux dans les résumés. On peut cependant s'arrêter sur des exemples qui révèlent, dans les motifs de jugement, la forte intrication des procès dans des structures syntaxiques complexes, caractérisées par l'accumulation de post-modifications entre le sujet et le prédicat principal, comme on le voit dans l'exemple (277).

- (277) Par conséquent, la conclusion du juge de première instance selon laquelle le délai ne **reposait** pas sur le postulat selon lequel les salariées **travaillant** dans des milieux où les femmes **sont** victimes de ségrégation **sont** "moins capables, ou moins dignes d'être **reconnu[e]s** ou **valorisé[e]s** en tant qu'être humain ou que membre de la société canadienne" (Law Canada (Ministre de l'Emploi et de l'Immigration), [1999] 1 S 497, par. 99) ne **peut aseptiser** l'effet discriminatoire (Québec A, par. 244, 327 et 330).
- (278) Le Québec **avance** [Procès : verb] que le par. 15(2) pourrait en l'espèce **s'appliquer** [Procès : rel] de manière à protéger l'art 38 d'une contestation fondée sur le par. 15(1). À mon avis, le par. 15(2) ne **s'applique** [Procès : rel] aucunement en l'espèce. (FR-CA-Jugements)

Dans l'exemple (277), le prédicat de la proposition principale *ne peut aseptiser* arrive après plus de cinquante mots, suite à l'accumulation de post-modifications du sujet *la conclusion*, non seulement par un GP mais également par une proposition relative qui contient elle-même un enchaînement de trois propositions relatives et une citation de la loi. Il s'agit d'un exemple, s'il en est, de *top-heavy sentence* qui affolerait bien des défenseur.euses du style clair, et qui illustre à nouveau le concept d'*intricacy*. On note également l'emploi d'un procès matériel métaphorique, *aseptiser* (qu'on trouve également dans la version anglaise du texte sous la forme de *sanitize*), qui n'apparaît ni dans les résumés, ni dans les textes législatifs (les métaphores étant déconseillées par les recommandations dans les textes de lois). Cet exemple rappelle que les juges du système de *common law* peuvent faire preuve d'un degré de personnalisation et d'un style imagé qui ne se verrait guère dans un arrêt de la Cour de Cassation française, et témoigne des différences de culture juridique entre droit civil et droit anglo-saxon (Galonnier 1997). Concernant les participants, comme en anglais, on remarque que les extraits de FR-CA-Jugements contiennent de participants très abstraits. Dans l'extrait (277), il s'agit de concepts comme *la conclusion*, *le délai*. Dans l'exemple (278), l'entité à l'origine du du procès verbal est à nouveau une entité collective abstraite (par métonymie), *le Québec*, et le procès relationnel *s'appliquera* pour sujet un GN qui réfère à un article de la loi. Comme ce que nous avons vu pour le passif, ce sont donc souvent des textes de lois ou des concepts et entités (très) abstraits qui sont en position thématique.

Nous pouvons comparer ces extraits, caractérisés par une syntaxe particulièrement complexes et des participants particulièrement abstraits, aux passages équivalents dans les résumés.

- (279) La juge Rosalie Silberman Abella, au nom des juges majoritaires de la Cour sur ce point, a **conclu** [Procès : verb] à la violation des droits des femmes à l'égalité. Selon elle, le délai **s'appliquait** [Procès : rel] à un groupe qui avait **fait** [Prédicat] l'objet [Procès : mat] d'une discrimination systémique **fondée** sur l'idée que le "travail des femmes" a une valeur moindre que le "travail des hommes". (FR-CA-Résumés)
- (280) Avec le concours de trois autres juges, la juge Suzanne Côté a **estimé** [Procès : men] qu'il n'y avait pas **eu** [Procès : rel] violation des droits des femmes à l'égalité. (FR-CA-Résumés)
- (281) La loi **perpétuait** [Procès : mat] ainsi le désavantage que **subissaient** [Procès : mat] déjà les femmes sur le marché du travail. (FR-CA-Résumés)

L'exemple (279) représente la simplification de la très longue phrase présentée dans l'extrait (277). La phrase ré-introduit un référent humain, *La juge Rosalie Silberman Abella* et un procès verbal (ce qui rappelle nos analyses des résultats quantitatifs). La phrase est également coupée en deux et contient principalement des procès matériels, dont un prédicat complexe lexicalisé *fait l'objet de*. Il est intéressant de noter que le sens de "faire l'objet", composé de faire + objet + de + N, dépend entièrement du sens du nom contenu dans le syntagme prépositionnel. La construction dans son ensemble, "avait fait l'objet d'une discrimination systémique" est équivalente ici, dans un style plus clair, au groupe verbal "a été discriminé". On voit à nouveau un exemple de procès ambigu, puisque "discriminé" peut être considéré comme un procès matériel (dans le sens de "a été séparé, distingué des autres") mais aussi comme un procès verbal (dans le sens de "insulté, offensé, outragé") (Gledhill 2008). On note également après le premier procès verbal que c'est un nom prédicatif qui est utilisé dans "a conclu à la violation des droits", et non pas une proposition "a conclu que les droits des femmes avaient été violés". Cette nominalisation déverbale se retrouve d'ailleurs dans l'exemple (280). Cet extrait suggère donc qu'une modification complète de la présentation de l'information a eu lieu, pour simplifier la syntaxe de la phrase et expliquer le raisonnement de la juge, mais que des noms prédicatifs et prédicat complexe ne disparaissent pas totalement des textes destinés au grand public. Dans les extraits (280) et (281), on voit cependant que ce sont davantage les acteurs institutionnels qui sont sujets et Agents, en particulier la figure du juge. Parfois, on trouve des agents abstraits comme *la loi*, mais il sont alors moins précis et plus généraux que la référence précise à l'article de loi dans l'extrait de jugement (279). Cette généralisation est due au rôle spécifique de

ces résumés, qui visent à synthétiser les arrêts mais aussi à expliquer les décisions qui font jurisprudence et ont des conséquences sur le droit fédéral, s'appliquant à tous.tes les justiciables. On voit dans ces extraits que la reformulation du discours des juges passe par une refonte totale de la structure des phrases, bien que l'influence du texte de départ soit toujours visible dans certaines formulations.

9.3 Analyse des procès sémantiques dans les textes de loi sur la discrimination et leur version en style clair

Nous passons à l'analyse des procès sémantiques dans les textes législatifs et leurs versions simplifiées, en anglais puis en français.

9.3.1 Entre EN-CAB et EN-Lois

Tout d'abord, nous rappelons que le Tableau 9.1 dans la Section 9.1.1. indique qu'il y a la même proportion de procès relationnels dans l'extrait de l'Equality Act et dans les extraits de EN-CAB. Comme nous l'avons déjà évoqué et le voyons dans les exemples (282) et (283), le genre législatif vise à définir les termes et la portée d'application du texte de loi lui-même. Les procès relationnels sont particulièrement prégnants dans cette échantillon, puisque la loi doit également définir les caractéristiques protégées contre la discrimination (comme la religion). C'est pourquoi on trouve dans l'extrait d'EN-Lois de nombreux exemples de procès relationnels et verbaux introduisant des définitions :

- (282) Religion **means** [Procès : ver] any religion and a reference to religion **includes** [Procès : rel] a reference to a lack of religion. (EN-Lois)
- (283) If the protected characteristic **is** [Procès : rel] marriage and civil partnership, this section **applies** [Procès : rel] to a contravention of B who **is** [Procès : rel] married or a civil partner. (EN-Lois)

On reconnaît dans ces exemples le caractère fortement méta-textuel des textes de lois, qui viennent expliciter les termes employés dans le texte lui-même et font référence à leur propre organisation textuelle (*this section*). Dans les textes rédigés en style clair, les procès relationnels tendent à recontextualiser l'information sur

la discrimination, de manière à ce que les justiciables puissent reconnaître si leur propre situation constitue cette infraction :

- (284) Racial segregation, if it's [Procès : rel] deliberate, is always **considered** [Procès : men] to **be** [Procès : rel] less favourable treatment. This **means** [Procès : ver] you can challenge [Procès : ver] racial segregation under the Equality Act as it's [Procès : rel] direct discrimination. (EN-CAB)
- (285) A youth club **has** [Procès : rel] different opening times for the local Asian and black communities. This **is** [Procès : rel] a deliberate policy by the youth club to avoid any trouble between the two communities. This **is** [Procès : rel] racial segregation and **is** [Procès : rel] unlawful discrimination.

Dans les exemples (284) et (285), on voit que les rédacteur.ices définissent la discrimination et l'illustrent par des procès relationnels, principalement sous la forme du verbe-copule *be*. Par le biais de mentaux et verbaux, ils expliquent également ce que la loi considère comme une discrimination (*considered*), et indiquent les procédures que peuvent lancer les justiciables (*challenge*). Parfois, des procès verbaux émanent d'une reformulation explicite, comme le segment métatextuel *This means* dans l'exemple (284), qui permet d'explicitier en quoi la ségrégation raciale est toujours une forme de discrimination directe.

En effet, l'un des types de procès qui est également très représenté dans les échantillons de EN-Lois et EN-CAB est celui des procès matériels, qui forment respectivement 37,9% et 30% des procès exprimés par les verbes codés. Cependant, comme pour les jugements, les types d'agents sont très différents entre les deux registres (spécialisé et simplifié). Nous analysons d'abord quelques extraits de l'échantillon d'EN-Lois.

- (286) A person A **discriminates** [Procès : mat] against a disabled person [Circ] if A **treats** [rédicat] B unfavourably [Procès : mat] because of something arising in consequence of B's disability. (EN-Lois)
- (287) A Minister of the Crown may [Prédicat] by regulations amend [Procès : mat] section 1 so as to add [Procès : mat] a public authority to the authorities that **are** [Prédicat] **subject** [Procès : mat] to the duty under subsection of that section. (EN-Lois)

On remarque d'abord dans l'exemple (286) deux prédicats complexes. D'une part, *is subject to*, équivalent de "faire l'objet" en français, et qui peut être paraphrasé par *is affected by*, qui introduit un autre nom prédicatif, *the duty*. D'autre part, *treats B unfavourably* est aussi un Prédicat complexe, où *favourably* constitue

un complément prédicatif obligatoire. Les participants restent cependant généralement abstraits et génériques, ce qui correspond à la fonction générique des textes législatifs. Par ailleurs, de nombreux procès matériels dans le texte de loi sont présentés sous la forme de passifs non-périphrastiques, ce qui fait écho aux résultats du chapitre 8. Dans les textes de EN-CAB, les référents humains peuvent parfois être abstraits, mais dans les situations décrites les procès matériels sont beaucoup plus concrets que *treat* ou *discriminate*, procès qui apparaissent respectivement 36 *treat* et 13 fois dans l'extrait de l'*Equality Act* étudié. À l'inverse, dans EN-CAB, on retrouve la stratégie de personnalisation et de concrétisation consistant à utiliser le pronom de deuxième personne et des référents humains concrets, associés à des procès matériels concrets, comme illustré dans les exemples (288) et (289).

- (288) Your employer **disciplines** [Procès : mat] you because you have to **take** [Procès : mat] time off to care for your disabled child. He has not **disciplined** [Procès : mat] other workers who have **had** similar amounts of time off work.
- (289) Direct discrimination is when you're **treated** [Procès : mat] differently and worse than someone else for certain reasons. The Equality Act **says** [Procès : verb] you've been **treated** [Procès : mat] less favourably.

Dans l'exemple (288), l'agent du procès matériel affectant *you* est un référent humain à la fois particulier et générique, dans une situation concrète très précise qui est décrite grâce à plusieurs procès matériels, pour illustrer un type de discrimination. Dans l'exemple (289), le procès est également matériel mais le passif est utilisé pour mettre les justiciables en premier dans les propositions subordonnées. Même lorsque les termes exacts de la loi sont repris, par exemple *less favourably*, on voit clairement l'effort de concrétisation et de re-personnalisation avec une paraphrase explicative précédent l'expression, et la thématization du pronom de deuxième personne *you*.

Enfin, le Tableau 9.1 suggère que l'échantillon d'EN-CAB contient deux fois plus de procès mentaux que l'extrait d'EN-Lois. Voici quelques exemples tirés d'EN-CAB :

- (290) A reference in section 21 or 22 or an applicable third requirement is to be **construed** [Procès : men] in accordance with this section. (EN-Lois)
- (291) All you need to **show** [Procès : ver] is that you were treated unfavourably because of something to do with your disability.
- (292) If you have **experienced** [Procès : men] discrimination, you can get [Procès : mat] help from the EASS discrimination helpline. (...) You can **find** [Procès : men] useful information about discrimination on the EHRC website at www.equalityhumanrights.com.

Les procès mentaux sont assez rares dans l'Equality Act, qui, nous l'avons vu, est composé de beaucoup de procès relationnels. Lorsqu'ils sont présents, ils sont la plupart du temps dans des prédicats au passif et sont des indications pour l'interprétation du texte par les juges, comme on le voit avec *construed* dans l'exemple (290). Dans EN-CAB, les procès mentaux sont principalement liés à deux usages. Soit les rédacteur.ices. expliquent aux justiciables comment établir la discrimination (comment la démontrer), par le biais d'un procès verbal démonstratif (sous-type de procès mental) comme *show* ; soit ils utilisent des procès mentaux pour parler de l'expérience de la discrimination pour les justiciables (voir *experienced* dans l'exemple (292)). Dans l'exemple (292), *find*, également un procès mental, permet renvoyer vers d'autres sites et pages spécialisées que les justiciables peuvent consulter, rappelant que ces textes sont des genres discursifs issus du Web et contiennent donc des liens hypertextes. Dans tous ces exemples, le medium (la personne qui fait l'expérience du procès mental) est réalisé par le pronom de deuxième personne, *you*. Ces exemples, et la plus grande proportion de procès mentaux par rapport au texte de loi, suggèrent que l'expérience intérieure (psychologique) de la discrimination est prise en compte par les rédacteur.ices pour les accompagner, suggérant à nouveau que les ressources grammaticales ont une dimension interpersonnelle dans EN-CAB.

9.3.2 Entre FR-Lois et FR-DILA

Nous comparons maintenant les extraits du Code pénal régissant la discrimination et l'incitation à la haine et à la discrimination avec les extraits de FR-DILA s'adressant aux justiciables. On trouve à peu près la même proportion de procès relationnels dans les deux extraits de corpus, car tout deux décrivent la discrimination à l'aide de procès relationnels d'identification. Le texte législatif, comme en anglais, précise les conditions d'application de la loi par des procès relationnels, en gras dans les exemples (293) et (294). La version simplifiée de ces définitions est présentée dans l'extrait de FR-DILA (295).

- (293) **Constitue** [Procès : rel] une discrimination toute distinction [Procès : men] opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille (...) (FR-Lois)
- (294) Les dispositions de l'article précédent ne **sont** [Procès : rel] pas applicables : Aux discriminations **fondées** [Procès : rel] sur l'état de santé, lorsqu'elles **consistent** [Procès : rel] en des opérations **ayant** [Prédicat] pour objet [Procès : rel] la prévention et la couverture du risque décès (FR-Lois).

- (295) Le fait de pousser [Procès : mat] des tiers à manifester [Procès : verb/mat] de la haine, de la violence ou de la discrimination à l'encontre de certaines personnes, en raison de leur religion ou de leur origine nationale ou ethnique, **constitue** [Procès : rel] une incitation à la haine raciale. (FR-DILA)

Dans l'extrait du Code Pénal, on note l'emploi du prédicat complexe *ayant pour objet* mais également plusieurs verbes exprimant typiquement des procès relationnels mais relativement peu fréquents, comme *consistent* et *constitue*. Ce dernier est également utilisé dans FR-DILA, comme on le voit dans l'exemple (295), ce qui démontre à nouveau l'influence du style juridique sur les textes destinés au grand public en français, là où les textes dans EN-CAB favorisent des verbes d'identification très simples et très fréquents comme *be*.

Concernant maintenant les procès verbaux, l'échantillon du corpus FR-Lois en contient environ 15%. Ils sont principalement méta-textuels et renvoient à d'autres passages du texte législatif (ex *sur un motif **mentionné** à l'article 225-1 du présent code*), ou au délit d'incitation à la haine déjà évoqué. Dans FR-DILA, les procès verbaux sont un peu plus rares et surtout plus concrets. Ils servent à expliquer, par exemple, ce que les justiciables ont le droit de dire ou de taire dans un contexte professionnel :

- (296) Une salariée doit **révéler** [Procès : verb] sa grossesse à son employeur uniquement avant de partir en congé maternité. (FR-DILA)

Surtout, on observe dans le Tableau 9.2 que les deux échantillons contiennent de nombreux procès matériels, notamment 40% dans FR-Lois. Comme pour l'anglais, on note dans ces échantillons une préférence pour des Agents abstraits et des tournures impersonnelles.

- (297) La discrimination **définie** [Procès : ver] aux articles 225-1 à 225-1-2, commise [Procès : mat] à l'égard d'une personne physique ou morale, est **punie** [Procès : mat] de trois ans d'emprisonnement et de 45 000 euros d'amende lorsqu'elle consiste [Procès : mat] : (...) 2) A **entraver** [Procès : mat] l'exercice normal d'une activité économique quelconque; 3) A refuser d'**embaucher** [Procès : mat], à **sanctionner** [Procès : mat] ou à **licencier** [Procès : mat] une personne (...) (FR-Lois)

Les procès matériels sont présents dans des structures qui rappellent une stratégie discursive déjà évoquée : l'utilisation du passif, puisqu'on a deux passifs non-

périphrastiques modifiant *discrimination* dans la première phrase. De même la peine est introduite par un passif sans Agent, ce qui, nous l'avons vu, est fréquent dans les textes législatifs car c'est l'état du sujet, et non l'Agent, qui importe. La dimension impersonnelle et la généralité propres aux textes de loi sont quant à elles réalisées dans cet exemple non par l'utilisation de lettres pour représenter les personnes comme en anglais, mais par l'infinitif : ainsi chacune des personnes auxquelles le texte s'applique peut instancier l'Agent. Si la tournure est extrêmement impersonnelle, l'article décrit cependant concrètement les actions qui constituent la discrimination, grâce aux procès exprimés sous forme nominale (*la fourniture*) ou verbale (*sanctionnée, licencier*). On remarque donc une plus grande précision qu'en anglais, où les termes *discriminate* et la prédicat complexe *treat someone unfavorably* restent plus vagues pour recouvrir une variété de situations et d'actions.

Dans les échantillons de FR-DILA, on voit une nette sur-représentation des procès matériels, avec plus de la moitié des verbes codés correspondant à ce type de procès. Cette fréquence importante s'explique car ces textes guident véritablement les justiciables en leur présentant les différents types de recours possibles en cas de discrimination, et comment ils se déroulent. Grâce à des procès matériels, les textes expliquent très concrètement aux lecteur.ices ce qu'ils peuvent faire en cas de situation de discrimination :

- (298) Une victime peut **utiliser** [Procès : mat] plusieurs types de recours en même temps. (FR-DILA)
- (299) **Composez** [Procès : mat] le 17 en cas d'urgence concernant un accident de la route, un trouble à l'ordre public ou une infraction pénale. (FR-DILA)
- (300) Vous pouvez **prendre** [Procès : mat] rendez-vous avec un délégué du Défenseur des droits. (...) Attention : **joindre** [Procès : mat] à votre courrier les photocopies des pièces relatives à votre saisine. (FR-DILA)
- (301) À savoir : les personnes **poursuivies** [Procès : mat] ne peuvent pas **être jugées** [Procès : mat] en comparution immédiate, en comparution sur reconnaissance préalable de culpabilité, ni en convocation par procès-verbal. (FR-DILA)

L'utilisation de l'impératif de deuxième personne (*Composez*) et de formes infinitives souligne la dimension fondamentalement procédurale de ces textes. Ces textes constituent véritablement une notice pour savoir quoi faire en cas de discrimination, selon les différents cas de figure rencontrés, que ce soit dans le cadre du travail, de la recherche d'un logement ou de la fonction publique. Les textes du corpus FR-DILA viennent donc expliciter pas à pas les étapes à suivre en s'appuyant notamment sur des procès matériels.

Concernant les procès mentaux, dans FR-lois, on en trouve 20%. Cela est notamment dû à l'évocation à plusieurs reprises de l'intention des législateur.ices via des segments méta-linguistique (ex *les conditions prévues par l'article 131-35*). Les textes contiennent également des précisions sur ce qui doit être démontré pour établir la discrimination, mais aussi des références explicites à la figure du juge :

- (302) Le juge **forme** [Prédicat] **sa conviction** [Procès : men] après avoir **ordonné** [Procès : verb], en cas de besoin, toutes les mesures d'instruction qu'il **estime** [Procès : men] utiles. (FR-Lois)

Ce passage contient plusieurs procès mentaux dont un prédicat complexe : le prédicat *forme* exprime un procès (matériel) mais c'est véritablement son COD qui exprime l'étendue du procès, *conviction* faisant ici référence à un procès mental, puisque la conviction formée par le juge est une idée. Nous revenons plus loin sur ce type de structures. Les procès verbaux (*ordonné*) et mentaux (*estime*) rappellent les pouvoirs liés au rôle des juges, qui peuvent dans le cas d'une infraction pénale prendre toutes les mesures nécessaires pour apporter des preuves afin de démontrer l'infraction, avant de mener une réflexion menant à une décision. Dans FR-DILA, les procès mentaux et verbaux sont davantage liés à l'expérience des justiciables.

- (303) Si vous avez **demandé** [Procès : verb] un logement dans le parc privé et que **estimez** [Procès : men] avoir été victime d'une discrimination à la location, vous pouvez exercer [Procès : mat] différents recours.
- (304) Vous devez **prouver** [Procès : verb] que votre non-sélection s'est faite uniquement en raison de critères discriminatoires.

Comme dans EN-CAB, la connaissance juridique est recontextualisée grâce aux pronoms de deuxième personne et des procès mentaux pour représenter ou anticiper l'expérience des justiciables (*vous estimez, vous devez prouver*). Cependant on remarque un goût pour les formes nominales, en particulier dans l'exemple (297), où c'est une nominalisation déverbale qui est utilisée (*votre non-sélection*) plutôt qu'une proposition (*vous n'avez pas été sélectionné.e*). Par ailleurs, nous souhaitons revenir aux exemples (300) et (301) cités plus haut. Ces exemples contiennent des mots ou expressions signalant le contenu important comme *Attention* ou *A savoir*, qui expriment des procès mentaux. Or, les textes de FR-DILA sont publiés en ligne, et contiennent des boutons et menus dépliant. Ces mots ou expressions constituent donc des marqueurs de cohésion discursive qui semblent anticiper une

lecture non-linéaire de ces textes, par exemple, on peut imaginer qu'un.e lecteur.ice commencerait la lecture en regardant d'abord les informations signalées comme importantes, ou en déroulant sur le menu "A savoir". Ce type de lecture non linéaire est favorisé par le lectorat des textes au format numérique (Askehave et Nielsen 2005), ce que les rédacteur.ices prévoient en balisant textuellement (et dans la structure de la page) le contenu important.

9.4 Transitivité et modalité dans le discours sur la discrimination : discussion des différences français-anglais

Après avoir commenté la distribution des procès sémantiques, nous passons à l'analyse de deux traits liés aux ressources transitives : les prédicats complexes et les auxiliaires de modalité.

9.4.1 Prédicats complexes dans le discours du droit avant et après simplification

Dans cette section, nous nous penchons sur les prédicats complexes. Dans les motifs de jugements en français et en anglais, dont l'un est la traduction de l'autre, les procès sémantiques sont à peu près distribués de la même manière, comme nous l'avons vu plus haut, avec une très faible différence entre procès relationnels (très légèrement supérieurs en anglais). Mais une différence intéressante apparaît entre les deux échantillons quant au nombre de prédicats complexes. En effet, l'extrait de jugement en français en contient près de deux fois plus que le texte en anglais. On trouve dans le texte français de nombreux prédicats complexes, dont nous présentons quelques exemples ci-dessous en comparant la version française et sa version en anglais :

- (305) Il serait également contraire à cet objectif de prétendre qu'une loi qui a [Prédicat] un effet [Procès : mat] discriminatoire sur les personnes auxquelles un régime est censé **venir [Prédicat] en aide [Procès : mat]** peut "tendre" ou être "nécessaire" à la réalisation de tout objet améliorateur (...). (FR-CA-Jugements)
- (306) And it would be equally inconsistent with the purpose to suggest that legislation that **has** [Prédicat] a discriminatory **impact** [Procès : mat] on a scheme's intended beneficiaries, can

"serve" or be "necessary to" any ameliorative purpose in the sense intended by this Court in Cunningham.(EN-CA-Jugements)

Plusieurs de ces prédicats complexes sont des expressions lexicalisées comme ci-dessus *venir en aide* ou *avoir un effet sur*, mais également des constructions qui ne sont pas lexicalisées, comme *tendre ou être nécessaire à la réalisation*, un prédicat complexe étant sans doute utilisé pour permettre de coordonner les deux GV et les post-modifier avec le même complément. Ce dernier exemple rappelle la citation de Bhatia (1994), qui explique que les prédicats complexes contenant des noms prédicatifs permettent aux acteurs du droit de présenter l'information de manière compacte et concise. D'autres prédicats complexes trouvés dans les motifs de jugements en français sont essentiellement des collocations spécialisées, comme ci-dessous *fait acception* (signifiant "préférer"), que l'on trouve dans une citation directe de la Charte des Droits et Libertés, texte central décrivant les caractéristiques protégées contre la discriminations au Canada, dont nous donnons également la traduction en anglais :

(307) La loi ne **fait** [Prédicat] **acception** de personne et s'applique également à tous (...). (FR-CA-Jugements)

(308) Every individual **is** equal before and under the law (...) (EN-CA-Jugements).

Dans les deux séries d'exemples, il y a pour chaque phrase un ou plusieurs prédicats complexes dans la version en français, mais le texte en anglais utilise soit un nom (*beneficiaries*, exemple (306)), soit un verbe seul avec une ré-orientation de la syntaxe de la phrase (*Every individual is equal*, exemple (308)), c'est-à-dire aucun ou un seul prédicat complexe en anglais. Or, rappelons ici que c'est la même juge qui a rédigé cette opinion dans les deux langues : cette différence semble donc liée à des choix linguistiques ou de traduction. La propension à utiliser des prédicats complexes serait-elle donc liée au français ? Lorsqu'on compare les textes de lois, l'extrait de *Equality Act* avec ceux du Code pénal français, on observe la situation inverse : le texte législatif anglophone contient presque deux fois plus de prédicats complexes que les articles de lois francophones. Dans l'échantillon d'EN-Lois, on trouve en effet des répétitions fréquentes des mêmes prédicats complexes, comme les suivants :

- (309) For the purposes of subsection X, a provision, criterion or practice is discriminatory (...) if- (...) it **puts**[Prédicat], or **would put** [Prédicat], persons with whom B shares the characteristic **at a particular disadvantage** [Procès : mat] when compared with persons with whom B does not share it, it puts, or **would put** [Prédicat], B at that **disadvantage** [Procès : mat] (...) (EN-Lois)
- (310) Where this Act imposes a duty to **make** [Prédicat] reasonable **adjustments** [Procès : mat] on a person, this section, sections 21 and 22 and the applicable Schedule apply. (...) A person who **is** [Prédicat] **subject** to a duty to **make** [Prédicat] **reasonable adjustments** [Procès : mat] is not(...) entitled to (...) (EN-Lois)

Ces exemples confirment à nouveau la tendance du style législatif anglophone à utiliser les formes nominales, puisque c'est un prédicat complexe qui est préféré aux verbes *disadvantage*, *adjust* ou *must make reasonable adjustments* à la place de *is subject to...*, ce dernier prédicat étant particulièrement complexe structurellement (d'autant qu'il introduit un autre prédicat complexe) et caractéristique du discours législatif. Nous suggérons cependant que la différence entre le français et l'anglais législatif provient d'une différence de taille entre les données, l'échantillon français comptant 187 verbes codés où se trouvent peu d'occurrences de ces constructions, contre presque 600 verbes codés pour l'extrait anglophone. Concernant le français, si l'on compare les prédicats complexes des textes juridiques entre eux, elles sont présentes dans les deux extraits de corpus avec à peu près la même fréquence, bien qu'ils soient légèrement plus représentés dans les motifs de jugement de la CSC (7,3 % contre 5,9% dans les articles de loi français). Nous suggérons donc que les prédicats complexes sont caractéristiques des discours juridiques anglophone comme francophone, et que les différences quantitatives proviennent d'une question d'échantillonnage. On trouve par exemple dans FR-Lois à plusieurs reprises la collocation *opérer une distinction*, fondamentale dans ce texte pour définir la discrimination. Cette collocation est bien un prédicat complexe, puisque le verbe est générique et c'est véritablement l'objet qui exprime le procès (mental).

Que deviennent ces constructions dans les textes du corpus PLAIN ? Dans les textes anglophones, les prédicats complexes sont plus rares comme on le voit dans le Tableau 9.1. EN-Brochures contient 2,9% de prédicats complexes sur l'ensemble des verbes codés, et seulement 1,3% dans EN-CA-Résumés. Ces textes semblent donc respecter les recommandations pour un style clair en évitant ce type de constructions, bien qu'on en trouve quelques occurrences notables :

- (311) If someone wants to justify discrimination, it will be more difficult to do so if they've failed to **make** [Prédicat] **reasonable adjustments** [Procès : mat] in relation to your disability.

- (312) You're being **put** [Prédicat] **at a disadvantage** [Procès : mat] because you need the dog to help you. (EN-CAB)

Ces exemples de prédicats complexes correspondent en fait à une réutilisation directe de la phraséologie de l'Equality Act, qui emploie ces expressions pour définir la discrimination (*put at a disadvantage*) et les moyens d'y remédier (*make reasonable adjustments*). Ces exemples montrent l'interaction constante entre le discours spécialisé d'origine et le discours destiné au grand public, ce dernier empruntant à la phraséologie juridique pour expliquer le droit, tout en recontextualisant l'information dans des situations auxquelles les non-spécialistes peuvent facilement s'identifier.

Les résumés des motifs de jugement anglophones contiennent également quelques exemples de prédicats complexes.

- (313) The trial judge found that the delay did not violate the women's equality rights. This was because the **distinction** [Procès : men] the government **made** [Prédicat] was not based on the women's sex (a ground of discrimination listed in the Canadian Charter of Rights and Freedoms) (EN-CA-Résumés)

Les rédacteur.ices s'inspirent ici de l'expression des juges, *draw a distinction*. On trouve également les nominalisations *was no violation* et *was no breach*, qui suggèrent là aussi l'influence du genre discursif de départ. Dans chaque cas, un verbe-support générique est utilisé (*make, draw, be*) mais c'est le COD qui exprime véritablement le procès, en particulier le procès mental *make a distinction* ou matériel *breach*.

Des quatre échantillons de corpus, c'est dans les trois textes de FR-DILA qu'on trouve la plus grande proportion de prédicats complexes. Cette différence est d'abord due à la présence de prédicats complexes lexicalisés, *porter plainte* (8 occurrences) et *mettre en place* (6 occurrences).

- (314) Toute personne victime d'incitation à la haine raciale peut **porter** [Prédicat] **plainte** [Procès : mat/verb] contre l'auteur du délit. (FR-DILA)
- (315) Le procureur peut aussi **mettre** [Prédicat] **en place** [Procès : mat] des mesures alternatives aux poursuites comme un stage de citoyenneté. (FR-DILA)

Le premier exemple, *porter plainte* est une collocation liée à la procédure judiciaire, et puisque les textes s'adressent à de potentielles victimes de discrimination, on comprend que ce prédicat complexe soit fréquente. Le verbe générique *porter*

sert de prédicat, mais c'est véritablement le nom *plainte* qui exprime le procès, que l'on peut analyser comme un procès matériel (puisqu'il entraîne l'ouverture d'une enquête la création de textes, par exemple de procès-verbaux,) mais aussi comme un procès verbal, puisque c'est par une action verbale que les justiciables réalisent le procès. Pour *mettre en place*, il s'agit d'un prédicat complexe lexicalisé de la langue générale, qui n'est pas lié au registre juridique. Comme pour l'anglais, la présence de prédicats complexes paraît également être due à l'influence du style juridique, puisqu'on retrouve dans certains extraits de FR-DILA l'équivalent de la construction *draw/make a distinction* :

- (316) Des **distinctions [Procès : men]** peuvent également être **faites [Prédicat]** entre femmes et hommes lors de la constitution des jurys de concours ou d'examens et lors de la désignation de ses représentants aux CAP et aux comités techniques, pour respecter la parité. (FR-DILA)

Dans les textes législatifs de départ, c'est le prédicat complexe *opérer une distinction* qui est utilisé pour définir la discrimination. Il apparaît que les textes, pour expliquer le droit, ne peuvent pas se détacher entièrement de la phraséologie juridique. Mais on remarque dans les deux cas, dans le passage de LEX à PLAIN, le prédicat est modifié, passant de *draw* à *make* en anglais et de *opérer* à *faire* en français. Les rédacteur.ices utilisent des verbes génériques plus fréquents, là où le texte juridique utilise un verbe d'un registre soutenu, moins courant. Dans les deux langues, il y a donc bien simplification, non pas de la structure en elle-même, mais d'un élément lexical (le pivot de cette construction). La fréquence de ces prédicats complexes est par ailleurs plus faible dans les textes de médiation juridique en anglais, en particulier par rapport à FR-DILA, ce qui laisse penser, à nouveau, que ce dernier sous-corpus reste plus proche du style et de la voix discursive de départ.

9.4.2 Modalité et procès sémantiques dans EN-CAB et FR-DILA

La modalité est un trait linguistique fréquemment cité dans les caractérisations du discours du droit. Nous avons déjà évoqué la question de la modalité en lien avec le passif, dans FR-DILA et EN-CAB c'est pourquoi nous nous sommes intéressée aux procès sémantiques en lien avec la modalité dans ces échantillons.

En effet, les Tableaux 9.1 et 9.2 révèlent un résultat surprenant concernant non pas la distribution des procès sémantiques en eux-mêmes des auxiliaires de modalité

dans les prédicats issus de FR-DILA et EN-CAB. Dans FR-DILA, 26,8 % des procès codés sont insérés dans des GV qui contiennent un auxiliaire de modalité, soit plus de deux fois plus que dans les textes extraits d'EN-CAB (12,6%). Or, dans EN-CAB, nous l'avons vu, les modaux sont associés au *hedging*. Ils sont donc conjoints à tous types de procès, mais apparaissent en particulier dans des phrases contenant des procès relationnels visant à indiquer si la situation en question constitue une discrimination (317), et à expliquer ce qui est une infraction ou pas selon la loi (318) :

- (317) This would probably **be** unlawful discrimination.
 (318) Discrimination because of something connected to your disability can sometimes **be** lawful.
 (319) There was enough information available to the pub employee for her to find out about your disability. You could therefore **argue** that she should have **known** about it.

Dans les deux premières occurrences, ces modaux sont bien liés à une précaution. Dans l'exemple (319), la modalité est simplement liée au mode irréel, puisque les rédacteur.ices décrivent des situations inventées pour donner un exemple de cas de discrimination liée au handicap. On peut également y voir des marqueurs de politesse dans des exemples comme "*you can find information*", invitant les justiciables à consulter d'autres sources. C'est donc une modalité liée à la possibilité, sans certitude, qui ressort. Qu'en est-il dans FR-DILA ? Dans ce corpus, ce sont les auxiliaires de modalité *pouvoir* et surtout *devoir* (22 occurrences) qui se détachent. *Il faut* est également utilisé 4 fois. L'analyse d'exemples révèle plusieurs emplois différents de ces auxiliaires :

- (320) (a) Aucun salarié ou agent public ne peut être **discriminé** au travail en matière d'embauche, de formation, de salaire....
 (b) Il peut **s'agir** de discrimination raciste, sexiste, homophobe, par l'âge ou selon l'état de santé.
 (321) La réception de la plainte ne peut pas vous **être refusée**.
 (322) Certains critères ne doivent pas **être retenus** pour choisir un locataire, car ils sont discriminatoires.
 (323) Vous devez **prouver** que votre non-sélection s'est faite uniquement en raison de critères discriminatoires.

Dans l'exemple (320a) et dans l'exemple (321), *pouvoir* est employé avec deux procès matériel (*discriminer*) et matériel/verbal (*refuser*), mais l'auxiliaire a dans

ces emplois une valeur déontique, d'obligation, pour rappeler leurs droits aux justiciables. En effet, la phrase *la réception de la plainte ne peut pas vous être refusée* peut être glosée par *on ne doit pas/on n'a pas le droit de vous refuser la plainte*, de même que l'exemple (320a) peut être paraphrasé par *personne ne doit discriminer quelqu'un*. On trouve d'ailleurs plus loin un exemple de la même phrase avec *devoir* : "Un agent public ne doit faire l'objet d'aucune discrimination." (FR-DILA). Dans l'exemple (320b), *il peut s'agir, pouvoir* est utilisé avec un procès relationnel, et on peut y voir ici une forme de *hedging*, puisque les rédacteur.ices suggère que la loi peut s'appliquer à la situation du justiciable ou pas, en fonction des situations individuelles. Dans les textes étudiés, la modalité est utilisée pour insister sur les droits, sur ce qu'il est possible de faire, (*Un salarié peut saisir le conseil des prud'hommes pour régler tout conflit sur un cas de discrimination.*) ou encore interdire (*aucun salarié ne peut être sanctionné ou licencié pour avoir relaté des faits de discrimination.*) d'après la loi contre la discrimination. Enfin, *devoir* est également utilisé pour rappeler les conditions nécessaires pour établir la discrimination avec un procès mental, *retenus*. Notre hypothèse est que les rédacteur.ices de FR-DILA construisent un ethos discursif (c'est-à-dire une voix discursive) qui vise à incarner l'autorité de la loi, pour rappeler ce que dit celle-ci, mais également pour indiquer aux justiciables ce que les institutions judiciaires peuvent attendre d'eux au cours d'une procédure. Cela confirme la différence avec le discours de médiation anglophone représenté dans EN-CAB. Là où les textes anglophones favorisent le *hedging* avec *may, can* et *might*, c'est la modalité déontique liée à l'obligation et à la loi qui semble l'emporter dans les extraits de FR-DILA. La voix discursive semble donc plus autoritaire que dans EN-CAB, puisqu'elle se veut le reflet du droit, la DILA étant une institution qui dépend directement du gouvernement. A l'inverse, l'organisme CAB adopte une approche davantage axée sur les justiciables, ce qui provient d'une culture plus *consumer-oriented* et démontre l'influence durable du PLM sur la communication entre les institutions et le grand public.

9.5 Résumé du Chapitre 9

Dans ce chapitre, nous avons effectué une étude de cas en nous intéressant à la distribution des procès sémantiques et aux ressources transitives dans des textes des corpus LEX et PLAIN portant sur la notion juridique de discrimination.

L'analyse révèle des divergences dans la distribution des types de procès, entre les types de textes juridiques et les sous-corpus issus du corpus PLAIN entre eux. Les textes de lois et CAB et DILA définissent des relations et actions matérielles entre les justiciables en lien avec la loi, d'où la plus grande proportion de procès relationnels et matériels. Les motifs de jugements contiennent une plus grande proportion de procès mentaux pour expliciter le raisonnement des juges, et leurs résumés contiennent davantage de procès verbaux pour rapporter explicitement leur parole. Dans les textes du corpus PLAIN, tous genres confondus, la distribution et les participants des procès sémantiques confirme la ré-orientation du discours de procès et sujets inanimés très abstraits à des référents humains, notamment les justiciables et acteurs du droit. L'étude des prédicats complexes et de la modalité, quant à elle, suggère d'une part que la phraséologie du droit imprègne en partie les textes destinés au grand public, et révèle également des différences dans le positionnement des rédacteur.ices qui reformulent ce que dit la loi entre les différents contextes culturels francophones et anglophones.

Conclusion

Le point de départ de cette thèse était une interrogation sur l'intelligibilité de textes juridiques, perçus comme hautement complexes. Partant de ce constat, nous avons étudié le passage de textes spécialisés à des textes de médiation juridique destinés à un lectorat non-expert au prisme du style clair. Nous avons utilisés pour cela deux corpus *ad hoc*, LEX et PLAIN, contenant chacun plusieurs genres discursifs rédigés en deux langues, l'anglais et le français, dans une perspective *genre-based* et *register-based*.

Concernant les caractéristiques observées du discours du droit, les analyses présentées sont en accord avec la littérature, puisqu'elles confirment plusieurs caractéristiques considérées typiques du *legalese*, notamment sa nature hautement formulaïque et la préférence pour des formes nominales pour exprimer les procès. En confrontant les recommandations du mouvement pour un style clair aux productions réelles d'institutions et organismes juridiques, nous avons également interrogé la notion de complexité telle qu'elle est formulée dans le mouvement du *plain language* et telle qu'elle est conceptualisée dans la recherche en linguistique. Nos résultats suggèrent en effet que les textes qui se réclament du style clair suivent en partie les recommandations présentes dans les guides rédactionnels, notamment en ce qui concerne la longueur des phrases, les scores de lisibilité et l'utilisation d'un vocabulaire plus courant, comme nous l'avons vu dans les Chapitres 6 et 7. L'approche *genre-based*, de même que certaines caractéristiques non-linguistiques de ces textes (mise en page, listes à puces) incitent cependant à nuancer ces résultats. Nous avons observé des résultats plus surprenants, dont la fréquence élevée de termes et de tournures passives au sein des textes PLAIN, respectivement dans le Chapitre 7 et le Chapitre 8. A nouveau, ces résultats concordent avec ceux obtenus dans des études précédentes, qui suggèrent que les règles du *plain language* et d'autres langues contrôlées ne sont que partiellement suivies. En particulier, les passifs, nombreux dans certains genres du corpus PLAIN, invitent à reconsidérer les recommandations et à les nuancer. La perception du passif comme intrinsèquement plus complexe que l'actif semble relever,

au regard des productions réelles, d'une idée reçue sur la complexité linguistique. L'analyse en contexte nous a permis d'identifier des phénomènes phraséologiques et de cohésion discursive, ainsi que des stratégies rhétoriques de la diffusion des connaissances spécialisées, qui expliquent en partie l'utilisation du passif sous différentes formes dans le corpus PLAIN, et confirment sa prépondérance dans le corpus LEX.

Dans ce travail, nous avons également partagé nos réflexions méthodologiques sur l'outillage de l'analyse de corpus, en proposant l'utilisation de méthodes classiques et de méthodes quantitatives empruntées aux statistiques et au TAL. Notre contribution porte notamment sur l'utilisation d'outils statistiques et d'apprentissage-machine en analyse de discours, en particulier les biais méthodologiques qu'ils peuvent entraîner de par les métriques utilisées. Les métriques de complexité décrites dans le Chapitre 6 semblent cohérentes avec les guides rédactionnels en PL, reflétant une même conception de la complexité linguistique. Il convient de signaler ici les difficultés rencontrées pour appliquer ces méthodes à nos corpus : d'une part des difficultés techniques liées à l'encodage des textes et à la compréhension des outils, mais également des difficultés méthodologiques (que nous avons déjà commentées), liées à un déséquilibre dans la constitution des corpus anglophones et francophones et à des biais dans les métriques elles-mêmes. Nous travail contribue ainsi à une réflexion méthodologique sur la tendance (relativement) récente de l'utilisation des outils du *machine-learning* en analyse de discours. En alliant ces méthodes quantitatives à des approches plus classiques de la linguistique de corpus, et à des analyses plus qualitatives et plus restreintes, nous avons proposé une synthèse de plusieurs approches, selon nous complémentaires. Nous espérons avoir démontré la pertinence du croisement de ces différentes méthodologies qui ont, à notre connaissance, été peu usitées dans le cadre de l'étude du style clair.

En ce qui concerne la dimension comparative français-anglais de notre travail, les résultats suggèrent des différences nettes entre le corpus DILA, issu d'une autorité française, avec les autres sous-corpus francophones et anglophones, qui sont issus de pays où le style clair a été érigé comme véritable stratégie de communication entre les institutions et les justiciables. Le corpus FR-DILA présente surtout des différences en ce qui concerne le positionnement des rédacteur.ices : la voix discursive construite présente une dimension interpersonnelle moindre, reflétant une position d'autorité, là où les sous-corpus anglophones semblent présenter une plus grande forme de solidarité avec le lectorat. Ces différences semblent partiellement liés à l'influence du PLM dans les pays anglo-saxons. Bien que les résultats soient

à prendre avec de fortes précautions, le Chapitre 9 suggère qu'il y a davantage de prédicats complexes en français dans les jugements de la CSC. Globalement, cependant, notre travail permet surtout de dégager des stratégies rhétoriques similaires à la fois dans l'expression du discours juridique et dans sa diffusion et reformulation en français et en anglais. Les similarités proviennent de stratégies de communication communes à toutes les langues lorsqu'il s'agit de recontextualiser un discours spécialisé ; concernant le corpus spécialisé, on peut y voir également les traces d'une influence historique durable du français sur l'anglais, et réciproquement de l'anglais sur le français dans les dernières décennies. Les divergences entre le français et l'anglais reflètent quant à elle à la fois des différences dans le système et la culture juridiques, et dans la culture dans la communication avec les non-spécialistes : l'influence de mouvements *consumer-oriented* comme le PLM est visible dans les textes anglophones ou issus à majorité anglophone. Ces travaux contribuent donc à l'étude de l'anglais de spécialité, dans une moindre mesure au français de spécialité, et plus généralement à la linguistique de corpus comparée.

Ce travail présente plusieurs limites dans la constitution des données et dans les analyses présentées, qui méritent d'être précisées. Concernant la constitution et l'exploitation des corpus, plusieurs dimensions ont été écartées. En particulier, la dimension diachronique dans certains sous-corpus (notamment dans EN- et FR-Lois, dont les textes ont été publiés à plusieurs années voire décennies d'écart), de même que certaines différences nationales (Royaume-Uni/Nouvelle-Zélande/Canada) n'ont pas été prises en compte volontairement, car une analyse détaillée de ces paramètres aurait dépassé le cadre temporel de cette étude. De même, certains problèmes d'échantillonnages dans les outils et les données, notamment pour l'étude lexicale des textes francophones ou dans le dernier chapitre d'analyses, empêchent de généraliser les conclusions observées, et invitent à reproduire ces études sur des échantillons plus larges de corpus. Par ailleurs, il est notable que certains aspects qu'il aurait été souhaitable d'étudier n'ont pas pu être couverts dans le cadre de cette thèse. En particulier, ce travail ne contient pas d'avis du lectorat, spécialiste ou non, sur les reformulations du discours juridique selon les règles du style clair ou non, alors que les études sur les langues contrôlées ou le style clair incluent souvent des études de ce type, afin d'observer si les usagers de ces textes préfèrent les versions en style clair ou non.

Notre travail, de par ces limites et manquements, ouvre donc de nombreuses perspectives de recherche. Pour rebondir sur la limite évoquée ci-dessus, l'un des prolongements envisagés de ce travail serait justement de réaliser une enquête visant

à recueillir l'avis de justiciables non-spécialistes, sur leur compréhension d'exemples tirés de nos corpus selon ces règles du style clair et d'autres contenant des formulations différentes. Il serait particulièrement intéressant de comparer le jugement de justiciables francophones de France sur des phrases reformulées et contenant davantage de marqueurs d'inter-personnalité. Concernant les expert.es, il serait judicieux de réaliser une enquête fondée sur un questionnaire leur demandant si le sens juridique (en droit) est équivalent dans des extraits de textes et leur version en style clair, puisque les juristes interrogent le bien-fondé et la pertinence de l'utilisation du style clair dans les textes spécialisés cette fois (Taylor 2018). Qu'il s'agisse par exemple de compléter ou d'améliorer les études présentées ici, ou de prendre en compte les dimensions diachroniques ou nationales de nos données, les corpus LEX et PLAIN présentés offrent de nombreuses autres perspectives d'analyse. Les mêmes approches méthodologiques pourraient également être appliquées à d'autres types de textes, comme des textes en FALC (Facile à Lire et Comprendre), afin d'étudier différents degrés de simplification et de continuer d'enrichir l'étude de la médiation juridique.

Bibliographie

- Adam, J.-M. (2011). *Les textes : types et prototypes : récit, description, argumentation, explication et dialogue*. Armand Colin.
- Albrespit, J. (2007). Atypical passives. *Études anglaises*, 60 (4) : 466–482.
- Anthony, F. (2018). Chief Justice Richard Wagner promises new era of transparency for Supreme Court. *The Globe and Mail*. *En ligne*.
- Anthony, L. (2014a). AntConc (version 3.4. 3). Tokyo, Japan : Waseda University.
- Anthony, L. (2014b). AntWordProfiler (version 1.4. 1). Tokyo : Waseda university.
- Askehave, I. and Nielsen, A. E. (2005). Digital genres : a challenge to traditional genre theory. *Information technology & people*.
- Asprey, M. and Rawson, C. (2005). *Clarity*. Routledge.
- Asprey, M. M. (2004). Plain Language around the world. *Plain Language for Lawyers, 3rd edn, Federation Press, Sydney*.
- Australian Office of Parliamentary Counsel (2016). *Plain English Manual*. Australian Government.
- Bailin, A. and Grafstein, A. (2001). The linguistic assumptions underlying readability formulae : A critique. *Language & communication*, 21 (3) : 285–301.
- Ballier, N., Gaillat, Thomas et Zarrouk, M., Simpkin, A., Bouyé, M., Sousa, A., and Stearns, B. (2019a). A franco-irish project for the automatic identification of criterial features in learners of English. In *EUROCALL Conference 2019*, page 100.
- Ballier, N., Gaillat, T., Simpkin, A., Stearns, B., Bouyé, M., and Zarrouk, M. (2019b). A supervised learning model for the automatic assessment of language levels based on learner errors. In *European Conference on Technology Enhanced Learning*, pages 308–320. Springer.

BIBLIOGRAPHIE

- Balmford, C. (2002). Plain language : beyond a movement. In *Fourth Biennial Conference of the PLAIN Language Association*.
- Banks, D. (2002). Systemic functional linguistics as a model for text analysis. *ASp. la revue du GERAS*, (35-36) : 23–34.
- Banks, D. (2008). *The development of scientific writing. Linguistic features and historical context*. Equinox.
- Banks, D. (2011). La linguistique systémique fonctionnelle : une approche sémantique et sociale. *Amadis*, (9) : 265–278.
- Banks, D. (2017). The extent to which the passive voice is used in the scientific journal article, 1985–2015. *Functional Linguistics*, 4 (1) : 1–17.
- Bansal, S. and Aggarwal, C. (2018). Textstat [python package]. Accessible sur <https://github.com/shivam5992/textstat>.
- Basaneže, K. D. (2017). Extended 11 binomial expressions in the language of contracts. *Phraseology in Legal and Institutional Settings : A Corpus-based Interdisciplinary Perspective*, page 203.
- Bawarshi, A. S. and Reiff, M. J. (2010). *Genre : An introduction to history, theory, research, and pedagogy*. Parlor Press West Lafayette, IN.
- Beaunoyer, E., Arsenault, M., Lomanowska, A. M., and Guitton, M. J. (2017). Understanding online health information : Evaluation, tools, and strategies. *Patient education and counseling*, 100 (2) : 183–189.
- Beedham, C. (1981). The passive in English, German and Russian. *Journal of Linguistics*, 17 (2) : 319–327.
- Beedham, C. (1987). The English passive as an aspect. *Word*, 38 (1) : 1–12.
- Benítez Castro, M. Á. et al. (2014). *Formal, syntactic, semantic and textual features of English shell nouns*. Universidad de Granada.
- Bhatia, V. K. (1983). Simplification v. Easification—the Case of Legal Texts. *Applied linguistics*, 4 (1) : 42–54.
- Bhatia, V. K. (1994). Cognitive structuring in legislative provisions. volume 1. London : Longman.

- Bhatia, V. K. (2004). *Worlds of written discourse : A genre-based view*. Bloomsbury Publishing.
- Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford handbook of linguistic analysis*.
- Biber, D., Gray, B., Staples, S., and Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research : Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46 : 100869.
- Biber, D. and Kurjian, J. (2007). Towards a taxonomy of web registers and text types : A multi-dimensional analysis. In *Corpus linguistics and the web*, pages 109–131. Brill Rodopi.
- Biel, Ł. (2004). Lexical bundles in EU law the impact of translation process on the patterning of legal language. *Corpus*, page 118.
- Biel, L. (2017). Lexical bundles in EU law : the impact of translation process on the patterning of legal language. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective.*, volume 10, page 26. London/New York : Routledge.
- Bisang, W. (2014). Overt and hidden complexity—two types of complexity and their implications. *Poznan Studies in Contemporary Linguistics*, 50 (2) : 127–143.
- Bock, B. (2018). Die Menschenrechte in „Leichter Sprache“. Vergleich der Modifikationsprinzipien von zwei Fassungen. pages 169–200.
- Bowker, L. and Pearson, J. (2002). *Working with specialized language : a practical guide to using corpora*. Routledge.
- Breeze, R. (2017). Giving voice to the law : Speech act verbs in legal academic writing. In *Phraseology in Legal and Institutional Settings*, pages 221–239. Routledge.
- Bureau de la traduction (2018). *Recommandations Langue claire et simple*. Travaux publics et Services gouvernementaux Canada.
- Cadou, E. (2014). *Introduction au droit. Leçon 1. Présentation générale du Droit*. Fondation Ius camp ; Politia-Académie des Sciences morales et politiques.
- Caffarel, A. (2009). La représentation grammaticale de l'expérience : transitivité et agence. *La linguistique systémique fonctionnelle et la langue française, Paris, L'Harmattan*, pages 67–88.

BIBLIOGRAPHIE

- Cao, D. (2007). *Translating law*. Multilingual Matters.
- Cardon, R. and Grabar, N. (2019). Détection automatique de phrases parallèles dans un corpus biomédical comparable technique/simplifié. In *TALN 2019-26ème Conférence sur le Traitement Automatique des Langues Naturelles*.
- Cassen, B. (2010). *Quelles langues pour la science ?* La Decouverte.
- Charrow, R. P. and Charrow, V. R. (1979). Making legal language understandable : A psycholinguistic study of jury instructions. *Columbia law review*, 79 (7) : 1306–1374.
- Clubic (2002). Lara, logiciel destiné aux rédacteurs administratifs, est disponible en ligne. *Actualités*.
- Collette, K., Benoît Barnet, M., Laporte, D., Pouëch, F., and Rui-Souchon, B. (2002). *Guide pratique de la rédaction administrative*. Ministère de la fonction publique.
- Commission, E. (2015). *English Style Guide. A Handbook for Authors and Translators in the European Commission*. Luxembourg : Office des publications de l'Union européenne.
- Commission Européenne (2011a). *How To Write Clearly*. Luxembourg : Office des publications de l'Union européenne.
- Commission Européenne (2011b). *Rédiger clairement*. Luxembourg : Office des publications de l'Union européenne.
- Compagnone, A. (2015). The reconceptualization of academic discourse as a professional practice in the digital age : A critical genre analysis of ted talks. *HERMES-Journal of Language and Communication in Business*, (54) : 49–69.
- Condamines, A. and Picton, A. (2014). Étude du fonctionnement des nominalisations déverbales dans un contexte de déspecialisation. *Congrès Mondial de linguistique Française*, pages 697–711.
- Conrad, S. and Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica (2004)*, 20 (2005) : 56–71.
- Conseil d'État (2017). *Guide de légistique*. SGG. SLQD. DQD.
- Coughlan, S. (2020). *Criminal Procedure 4th ed*. Irwin Law Inc.

- Coulthard, M., Johnson, A., and Wright, D. (2016). *An introduction to forensic linguistics : Language in evidence*. Routledge.
- Cour Suprême du Canada (2017). *Rôle de la Cour*. Site web de la Cour Suprême du Canada. URL :<https://scc-csc.ca/court-cour/role-fra.aspx>.
- Cour Suprême du Canada (2022). *La cause en bref*. Site web de la Cour Suprême du Canada. URL=<https://scc-csc.ca/case-dossier/cb/index-fra.aspx>.
- Crossley, S. (2020). Linguistic features in writing quality and development : An overview. *Journal of Writing Research*, 11 (3).
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., and Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers : new approaches to readability formulas. *Discourse Processes*, 54 (5-6) : 340–359.
- Crystal, D. and Davy, D. (1969). Investigating English style : English language series. *London and Harlow : Longmans*.
- Curtotti, M., McCreath, E., Bruce, T., Frug, S., Weibel, W., and Ceynowa, N. (2015). Machine learning for readability of legislative sentences. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 53–62.
- Cutts, M. (2008). Plain English lexicon. *London : Plain Language Commission*. Retrieved on June, 15 : 2008.
- Cutts, M. (2013). *Oxford guide to plain English*. Oxford : Oxford University Press.
- Dąbrowski, A. (2016). Reel justice in the context of teaching Legal English as a foreign language. *Glottodidactica. An International Journal of Applied Linguistics*, 43 (2) : 107–120.
- De Marneffe, M.-C. and Manning, C. D. (2008). *Stanford typed dependencies manual*. Rapport technique, Stanford University.
- Défenseur des Droits, L. (2019). Dématérialisation et inégalités d'accès aux services publics. *Rapport. République Française*. Repéré à www.defenseurdesdroits.fr.
- Deumier, P. (2019). *Introduction générale au droit*. LGDJ.
- Diani, G. (2018). Popularization of legal knowledge in English and Italian information books for children. In *Popularization and Knowledge Mediation in the Law*, pages 291–316. Münster : Lit Verlag.

BIBLIOGRAPHIE

- Diani, G. (2019). English and Italian land contracts : A cross-linguistic analysis. In *Corpus-based Research on Variation in English Legal Discourse*, pages 25–50. John Benjamins.
- Dieckmann, C. and Rojas-Lizana, I. (2016). The pragmatics of legal advice services in a community legal centre in australia : domination or facilitation ? *International Journal of Speech, Language & the Law*, 23 (2).
- Digital Government New Zealand (2020). *Plain language*. New Zealand Government Information Services. URL : <https://www.digital.govt.nz/standards-and-guidance/design-and-ux/content-design-and-management/plain-language/>.
- Eagleson, R. D., Jones, G., and Hassall, S. (1990). *Writing in Plain English*. Australian Government Pub. Service.
- Farrington, G. (1996). AECMA simplified English. An overview of the international aerospace maintenance language. *Proc. CLAW96, Katholieke Universiteit Leuven, Leuven*, pages 1–21.
- Ferraresi, A. (2019). How specialized (or popularized) ? Terminological density as a clue to text specialization in the domain of food safety. *Lingue e linguaggi*, 29.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32 : 221–223.
- Flesch, R. (2007). Flesch-kincaid readability test. *Retrieved October, 26 (3) : 2007*.
- Flesch, R. F. (1979). *How to write in plain English : A book for lawyers and consumers*. HarperCollins.
- Fløttum, K., Jonasson, K., and Norén, C. (2007). *On : pronom à facettes*. De Boeck/Duculot.
- Foudon, N. (2008). *Acquisition du langage chez les enfants autistes : Etude longitudinale*. Thèse de doctorat, Université Lyon 2.
- François, T., Brouwers, L., Naets, H., and Fairon, C. (2014). Amesure : a readability formula for administrative texts (amesure : une plateforme de lisibilité pour les textes administratifs)[in french]. In *Proceedings of TALN 2014 (Volume 2 : Short Papers)*, pages 467–472.
- François, T. and Fairon, C. (2013). Les apports du TAL à la lisibilité du français langue étrangère. *Traitement Automatique des Langues*, 54 (1).

- François, T., Müller, A., Rolin, E., and Norré, M. (2020). Amesure : a web platform to assist the clear writing of administrative texts. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing : System Demonstrations*, pages 1–7.
- Gaillat, T., Ballier, N., Sousa, A., Bouyé, M., Simpkin, A., Stearns, B., and Zarrouk, M. (2020). Un prototype en ligne pour la prédiction du niveau de compétence en anglais des productions écrites (a prototype for web-based prediction of English proficiency levels in writings). In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4 : Démonstrations et résumés d'articles internationaux*, pages 30–33.
- Gaillat, T., Simpkin, A., Ballier, N., Stearns, B., Sousa, A., Bouyé, M., and Zarrouk, M. (2022). Predicting cefr levels in learners of english : the use of microsystem criterial features in a machine learning approach. *ReCALL*, 34 (2) : 130–146.
- Galonnier, B. (1997). Le discours juridique en France et en Angleterre. Convergences et spécificités. *Asp. la revue du GERAS*, (15-18) : 427–438.
- Gangloff, D. (2015). Le langage facile à lire et à comprendre. *Revue suisse de pédagogie spécialisée*, 4 : 44–50.
- Gardner, D. and Davies, M. (2014). A new academic vocabulary list. *Applied linguistics*, 35 (3) : 305–327.
- Gautier, L. (2003). Terminologie et phraséologie comparées du droit constitutionnel en français et en allemand. In *L'espace euro-méditerranéen : une idiomaticité partagée*, number 9973-902-28-9, pages Pages–113. Centre d'Etudes et de Recherches Economiques et Sociales.
- Gémar, J.-C. (1981). Réflexions sur le langage du droit : problèmes de langue et de style. *Meta*, 26 (4) : 338–349.
- George, G. F. A. (2010). The Woolf reform of civil procedure : a possible end to legalese ? *LSP Journal-Language for special purposes, professional communication, knowledge management and cognition*, 1 (2).
- Gibbons, J. (1995). Language and the Law. volume 73, page 285. Wiley Online Library.

BIBLIOGRAPHIE

- Gibbons, J. (1999). Language and the law. *Annual Review of Applied Linguistics*, 19 : 156–173.
- Giry-Schneider, J. (1987). *Les prédicats nominaux en français : les phrases simples à verbe support*, volume 18. Librairie Droz.
- Gledhill, C. (2008). *Le signe et le syntagme : entre phraséologie et lexico-grammaire. Une synthèse du modèle systémique fonctionnel de Michael Halliday et de la théorie sémiotique de Charles S. Peirce*. Thèse d'habilitation. Université de Bretagne occidentale.
- Gledhill, C. (2011). Simplified technical English : from the description of a new technolect to a critique of new new-speak.
- Gledhill, C. and Kübler, N. (2016). What can linguistic approaches bring to English for Specific Purposes ? *ASp. la revue du GERAS*, (69) : 65–95.
- Gledhill, C., Martikainen, H., Mestivier, A., and Zimina-Poirot, M. (2019). Towards a linguistic definition of 'simplified medical English' : Applying textometric analysis to cochrane medical abstracts and their plain language versions. *LCM-La Collana/The Series*, pages 91–114.
- Gledhill, C., Patin, S., and Zimina, M. (2017). Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français. *Corpus*, (17).
- Gobbo, F. (2017). Are planned languages less complex than natural languages ? *Language Sciences*, 60 : 36–52.
- Goldberg, A. E. (2006). *Constructions at work : The nature of generalization in language*. Oxford University Press on Demand.
- Government Digital Service (2016). *Writing for GOV.UK*. GOV.UK. En ligne. Consulté le : 02/02/2018.
- Goźdź Roszkowski, S. and Pontrandolfo, G. (2013). Evaluative patterns in judicial discourse : a corpus-based phraseological perspective on american and italian criminal judgments.
- Goźdź-Roszkowski, S. and Pontrandolfo, G. (2017). *Phraseology in legal and institutional settings : a corpus-based interdisciplinary perspective*. Routledge.

- Greene, E., Fogler, K., and Gibson, S. C. (2012). Do people comprehend legal language in wills? *Applied cognitive psychology*, 26 (4) : 500–507.
- Grice, H. P. (1975). Logic and conversation. In *Speech acts*, pages 41–58. Brill.
- Groom, N. (2019). Construction grammar and the corpus-based analysis of discourses : The case of the WAY IN WHICH construction. *International Journal of Corpus Linguistics*, 24 (3) : 291–323.
- Gustafsson, M. (1984). The syntactic features of binomial expressions in legal English. *Text-Interdisciplinary Journal for the Study of Discourse*, 4 (1-3) : 123–142.
- Halliday, M. A. K. and Matthiessen, C. M. (2013). *Halliday's introduction to functional grammar*. Routledge.
- Halliday, M. A. K. and Webster, J. J. (2009). *Bloomsbury Companion to Systemic Functional Linguistics*. A&C Black.
- Hansen-Schirra, S., Bisang, W., Nagels, A., Gutermuth, S., Fuchs, J., Borghardt, L., Deilen, S., Gros, A.-K., Schiff, L., and Sommer, J. (2020). Intralingual translation into easy language—or how to reduce cognitive processing costs. *Easy Language Research : Text and User Perspectives*. Berlin : Frank & Timme, pages 197–225.
- Hartig, A. J. and Lu, X. (2014). Plain English and legal writing : Comparing expert and novice writers. *English for Specific Purposes*, 33 : 87–96.
- Harvey, K., Locher, M. A., and Mullany, L. (2013). " can I Be at Risk of Getting AIDS ?" A Linguistic Analysis of Two Internet Columns on Sexual Health. *Linguistik online*, 59 : 111–132.
- Hashim, N. A. (2013). Plain Language : give it a try ! *Commonwealth Law Bulletin*, 39 (3) : 423–434.
- Hilton, H. (2002). Modèles de l'acquisition lexicale en l2 : où en sommes-nous ? *ASp. la revue du GERAS*, (35-36) : 201–217.
- Huddleston, R. and Pullum, G. (2005). The cambridge grammar of the English language. *Zeitschrift für Anglistik und Amerikanistik*, 53 (2) : 193–194.
- Hunston, S. and Francis, G. (2000). *Pattern grammar : A corpus-driven approach to the lexical grammar of English*. Number 4. Studies in Corpus Linguistics, John Benjamins Publishing.

BIBLIOGRAPHIE

- Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *Tesol Quarterly*, pages 195–202.
- Hupet, M. and Costermans, J. (1976). Un passif : Pour quoi faire ?, quinze années de travaux psycholinguistiques. *La linguistique*, 12 (Fasc. 2) : 3–26.
- Jacobi, D. (1985). Sémiotique du discours de vulgarisation scientifique. *Semen. Revue de sémio-linguistique des textes et discours*, (2) : En ligne.
- Jelicic Kadic, A., Fidahic, M., Vujcic, M., Saric, F., Propadalo, I., Marelja, I., Dosenovic, S., and Puljak, L. (2016). Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. *BMC medical research methodology*, 16 (1) : 1–4.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing : A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 53 : 61–79.
- Kerbrat-Orecchioni, C. (1991). La politesse dans les interactions verbales. *Dialoganalyse III*, 1 : 39–59.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., and Suchomel, V. (2014). The Sketch Engine : ten years on. *Lexicography*, 1 (1) : 7–36.
- Koester, A. (2010). Building small specialised corpora. In *The Routledge handbook of corpus linguistics*, pages 94–107. Routledge.
- Kraif, O. and Tutin, A. (2017). Des motifs séquentiels aux motifs hiérarchiques : l'apport des arbres lexico-syntaxiques récurrents pour le repérage des routines discursives. *Corpus*, 6 (17).
- Lam, T., Demange, J., and Longhi, J. (2021). Attribution d'auteur par utilisation des méthodes d'apprentissage profond. In *EGC 2021 Atelier" DL for NLP : Deep Learning pour le traitement automatique des langues"*.
- Laufer, B. and Nation, P. (1995). Vocabulary size and use : Lexical richness in L2 written production. *Applied linguistics*, 16 (3) : 307–322.
- Leete, B. A. (1981). plain language legislation : A comparison of approaches. *American business law journal*, 18 (4) : 511–524.

- Lehmann, D. (1997). English for specific purposes (ESP) et français sur objectifs spécifiques (FOS) : le préalable du contenu préalable. *ASp. la revue du GERAS*, (15-18) : 11–26.
- Lehto, A., Fanego, T., and Rodríguez-Puente, P. (2019). The representation of citizens and monarchy in Acts of Parliament in 1800 to 2000. *Corpus-based research on variation in English legal discourse*, pages 235–260.
- Longhi, J. (2019). Contours, perspectives et tensions des «humanités numériques». *Sens-Dessous*, (2) : 123–135.
- Lopez, S., Condamines, A., Josselin-Leray, A., O'Donoghue, M., and Salmon, R. (2013). Linguistic Analysis of English Phraseology and Plain Language in Air-ground Communication. *halshs.archives-ouvertes.fr*.
- LoPucki, L. M. (2014). The Readable Delaware General Corporation Law 2014-2015 with Patented VisiLaw Markings. *UCLA School of Law Research Paper*, (14-15).
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15 (4) : 474–496.
- Lu, X. (2012). The relationship of lexical richness to the quality of esl learners' oral narratives. *The Modern Language Journal*, 96 (2) : 190–208.
- Lu, X. and Ai, H. (2015). Syntactic complexity in college-level English writing : Differences among writers with diverse l1 backgrounds. *Journal of Second Language Writing*, 29 : 16–27.
- MacGeorge, E. and Van Swol, L. (2019). Advice : Communicating to support and influence. In *Oxford Research Encyclopedia of Communication*.
- Maher, T. (2021). The a to z alternative. *Communication personnelle (courriel)*.
- Maingueneau, D. (2013). Genres de discours et web : existe-t-il des genres web? *Manuel d'analyse du web en Sciences Humaines et Sociales*, page 72.
- Maingueneau, D. (2016). *Les termes clés de l'analyse du discours*. Le seuil.
- Maley, Y. (1994). The language of the law. volume 1, pages 11–50. London : Longman.
- Masson, M. E. and Waldron, M. A. (1994). Comprehension of legal contracts by non-experts : Effectiveness of plain language redrafting. *Applied Cognitive Psychology*, 8 (1) : 67–85.

BIBLIOGRAPHIE

- Matoesian, G. M. (1997). “You Were Interested in Him as a Person?” : Rhythms of Domination in the Kennedy Smith Rape Trial. *Law & Social Inquiry*, 22 (1) : 55–93.
- Mazzi, D. (2017). “By partially renouncing their sovereignty...” : on the discourse function(s) of lexical bundles in EU-related Irish judicial discourse. In *Phraseology in Legal and Institutional Settings. A Corpus-Based Interdisciplinary Perspective*, pages 189–202. Routledge.
- McLaughlin, G. (1969). Smog grading – a new readability formula.
- Minton, T. D. (2015). In defense of the passive voice in medical writing. *The Keio Journal of Medicine*, 64 (1) : 1–10.
- Monjean-Decaudin, S. (2010). Territorialité et extraterritorialité de la traduction du droit. *Meta : Journal des traducteurs/Meta : Translators’ Journal*, 55 (4) : 693–711.
- Nelson, M. (2010). Building a written corpus : What are the basics ? In *The Routledge handbook of corpus linguistics*, pages 81–93. Routledge.
- Neveu, F. (2004). *Dictionnaire des sciences du langage*. Armand Colin.
- Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., et al. (2016). Universal dependencies v1 : A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 1659–1666.
- Office of the Scottish Parliamentary Counsel (2006). *Plain Language and Legislation. Drafting Matters! Guidance*. Scottish Government.
- Ogden, C. K. and Halász, G. (1935). *Basic English*. Kegan Paul Trench Trubner.
- Orlando, D. (2018). The problem of legal phraseology : a case of translators vs lawyers.
- Orts Llopis, M. (2009). Legal genres in English and spanish : some attempts of analysis. *Ibérica*, (18).
- Orwell, G. (1946). *Politics and the English language*.

- Oursel, E. (2012). Le refus de service dans les administrations : quelles formes d'atténuation pour ces actes menaçants et quelles offres de compensation ? *Signes, Discours et Sociétés : Revue semestrielle en sciences humaines et sociales dédiée à l'analyse des Discours*, 9 : En-ligne.
- O'Brien, S. (2010). Controlled language and readability. volume 15, pages 143–168. John Benjamins Amsterdam.
- Parliamentary Counsel Office (2021). *Principles of clear drafting*. New Zealand Government.
- Pati, S., Kavanagh, J. E., Bhatt, S. K., Wong, A. T., Noonan, K., and Cnaan, A. (2012). Reading level of Medicaid renewal applications. *Academic Pediatrics*, 12 (4) : 297–301.
- Petelin, R. (2010). Considering plain language : issues and initiatives. *Corporate Communications : An International Journal*.
- Peterson, D. A. (1990). Developing a Simplified English Vocabulary. *Technical communication*, 37 (2) : 130–33.
- Pic, E. and Furmaniak, G. (2010). Grammaire et degré de spécialisation. *ASp. la revue du GERAS*, (58) : 39–55.
- Pic, E. and Furmaniak, G. (2014). Impact du lectorat visé sur la grammaire. bilan du projet «sensibilité de la grammaire anglaise au degré de spécialisation». *ASp. la revue du GERAS*, (65) : 69–86.
- Pishghadam, R. and Abbasnejad, H. (2016). Emotioncy : A potential measure of readability. *International Electronic Journal of Elementary Education*, 9 (1) : 109–123.
- PLAIN (2011). *Federal plain language guidelines*. Plain Language Action and Information Network. URL : <http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/>.
- Plain English Campaign (2001a). *The A to Z of alternative words*. Plain English Campaign Drafting Resources. URL : www.plainenglish.co.uk/free-guides.html.
- Plain English Campaign (2001b). *How to write medical information in Plain English*. Plain English Campaign Drafting Resources. URL : www.plainenglish.co.uk/free-guides.html.

BIBLIOGRAPHIE

- Plain English Campaign (2022). *How to write in Plain English*. Plain English Campaign Drafting Resources. URL : www.plainenglish.co.uk/free-guides.html.
- Plain Language Association International (2022). *What is plain language ?* Plain Language Association International (PLAIN). URL : <https://plainlanguagenetwork.org/plain-language/what-is-plain-language/>
Consulté le : 5/10/2017.
- Preite, C. (2016). La vulgarisation des termes juridiques et la construction d'un savoir («que» faire) chez le grand public. *REPÈRES-DORIF*, 10 : 1–9.
- Preite, C. (2018). Stratégies dialogiques et transmission du savoir juridique dans le site du ministère de la justice française. *Popularization and Knowledge Mediation in the Law. Popularisierung und Wissensvermittlung im Recht*, 9 : 149.
- Preite, C. et al. (2007). La politesse linguistique dans les arrêts de la cour de justice des communautés européennes. *Synergies Italie*, 3 : 94–101.
- Raus, R. (2014). *La terminologie multilingue : la traduction des termes de l'égalité H/F dans le discours international*. De Boeck Supérieur.
- Richard, I. (2006). L'évolution de l'emploi de shall, de must et du présent simple dans le discours juridique normatif dans le cadre du plain language movement. *ASp. la revue du GERAS*, 68 (49-50) : 137–153.
- Richard, I. (2018). Is legal lexis a characteristic of legal language? *Lexis. Journal in English Lexicology*, (11).
- Richaudeau, F. (1976). Faut-il brûler les formules de lisibilité? *Communication & Langages*, 30 (1) : 6–19.
- Rossetti, A., Cadwell, P., and O'Brien, S. (2020). "the Terms and Conditions Came Back to Bite" : Plain Language and Online Financial Content for Older Adults. In *International Conference on Human-Computer Interaction*, pages 699–711. Springer.
- Rouski, M. (2015). Le passif de l'autorité. *Corela. Cognition, représentation, langage*, (13-1).
- Schmid, H. (1994). Treetagger. *Proceedings of International Conference on New Methods in Language Processing*.

- Searle, J. R. (1965). What Is A Speech Act? *Perspectives in the Philosophy of Language : a Concise Anthology*, 2000 : 253–268.
- Seoane, E., Williams, C., Dossena, M., and Taavitsainen, I. (2006). Changing the Rules : A Comparison of Recent Trends in English in Academic Scientific Discourse and Prescriptive Legal Discourse. In *Diachronic Perspectives on Domain-specific English*, volume 40 of *Studies in Language and Communication*, page 255–271. Peter Lang, linguistic insights edition.
- Simpkin, T. G. N. B. A., Stearns, B., and Zarrouk, M. B. M. (2019). Investigating criterial features of learner English : Microsystem. *Workshop. Beyond CEFR level prediction of texts in learner corpora : Exploring feedback to learners and learning analytics*.
- Skibitska, O. (2015). The language of tourism : Translating terms in tourist texts. *Translation Journal*, 18 (4).
- Sousa, A., Ballier, N., Gaillat, T., Stearns, B., Zarrouk, M., Simpkin, A., and Bouyé, M. (2020). From linguistic research projects to language technology platforms : A case study in learner data. In *Proceedings of the 1st International Workshop on Language Technology Platforms*, pages 112–120.
- Steadman, J. (2013). *Drafting legal documents in plain English - Guida pratica alla redazione di documenti legali in inglese. Ediz. bilingue*. Giuffrè Editore.
- Stroup, D. G. (1983). Law and language : Cardozo's jurisprudence and Wittgenstein's philosophy. *Val. UL Rev.*, 18 : 331.
- Szczyrbak, M. (2017). Verba dicendi in courtroom interaction : Patterns with the progressive. In *Phraseology in legal and institutional settings*, pages 240–257. Routledge.
- Taylor, S. (2003). La rédaction et l'interprétation des textes législatifs français et anglais : une convergence des cultures juridiques. *Greenstein R., Langues, cultures et codes : regard croisés, Paris, L'Harmattan*, pages 99–117.
- Taylor, S. (2018). UK Sale of Goods Legislation 1893-2015 : Towards Plain(er) Language? *ASp. la revue du GERAS*, (74) : 95–112.
- Thrush, E. A. (2001). Plain English? A study of plain English vocabulary and international audiences. *Technical communication*, 48 (3) : 289–296.
- Tiersma, P. M. (2000). *Legal language*. University of Chicago Press.

- Trklja, A. (2017). A corpus investigation of formulaicity and hybridity in legal language : a case of eu case law texts. In *Phraseology in Legal and Institutional Settings*, pages 89–108. Routledge.
- Turfler, S. (2015). Language ideology and the plain-language movement : How straight-talkers sell linguistic myths. *Legal Comm. & Rhetoric : JAWLD*, 12 : 195.
- Turnbull, J. (2014). Expert To Lay Communication : Legal Information and Advice on the Internet. *Language and Law in Professional Discourse : Issues and Perspectives*, page 61.
- Turnbull, J. (2018). Communicating and recontextualizing legal advice online in English. *Popularization and knowledge mediation in the law. Popularisierung und Wissensvermittlung im Recht*, pages 201–222.
- Ure, J. (1971). Lexical density and register differentiation. In *Selected Papers of the Second International Congress of Applied Linguistics. Applications of Linguistics*. London : Cambridge University Press.
- US Department of Transportation (2014). *Writing Reader-Friendly Documents*. CreateSpace Independent Publishing Platform.
- Vajjala, S. and Meurers, D. (2014). Readability assessment for text simplification : From analysing documents to identifying sentential simplifications. *ITL-International Journal of Applied Linguistics*, 165 (2) : 194–222.
- Warnier, M. (2018). *Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux*. Thèse de doctorat, Université Toulouse Jean-Jaurès.
- Williams, C. (2004). Legal English and plain language : An introduction. *ESP Across Cultures*, 1 (1) : 111–124.
- Williams, C. (2013). Changes in the verb phrase in legislative language in English. *The verb phrase in English : Investigating recent language change with corpora*, page 353.
- Williams, C. (2015). Changing with the Times : The Evolution of Plain Language in the Legal Sphere. *Revista Alicantina de Estudios Ingleses.*, 28 : 183.
- Williams, C. (2020). Review of fanego, teresa and paula rodríguez-puente eds. 2019. corpus-based research on variation in English legal discourse. amsterdam : John

- benjamins. isbn : 978-9-027-20235-2. <https://doi.org/10.1075/scl.91>. *Research in Corpus Linguistics*, 8 (1) : 178–194.
- Williams, J. and Bizup, J. (2017). *Style : Lessons in Clarity and Grace*. Always learning. Pearson.
- Xanthaki, H. (2019). Courts and legislation : Do legislators and judges speak the same language? In *Courts, Politics and Constitutional Law*, pages 57–68. Routledge.
- Yeught, M. V. d. (2016). Protocole de description des langues de spécialité. *Recherche et pratiques pédagogiques en langues de spécialité. Cahiers de l'Aplut*, 35 (spécial 1).
- You, T. (2019). Review : Stanisław goźdz-roszkowski and Gianluca Pontrandolfo : Phraseology in Legal and Institutional Settings : A Corpus-Based Interdisciplinary Perspective.

Annexes

Annexe A

Liste des métriques utilisées pour la classification

Métriques calculées sur les textes en anglais et en français

Le tableau ci-dessous présente les métriques communes aux jeu de données anglais et français.

Métrique	Outil de mesure	Dimension de la complexité mesurée	Description
W	L2SCA	Syntaxique	Nombre de mots par texte
S	L2SCA	Syntaxique	Nombre de phrases par texte
VP	L2SCA	Syntaxique	Nombre de groupes verbaux
C	L2SCA	Syntaxique	Nombre de propositions par texte
T	L2SCA	Syntaxique	Nombre de T-units par texte
CP	L2SCA	Syntaxique	Nombre de coordinations par texte
MLS	L2SCA	Syntaxique	Longueur moyenne des phrases

ANNEXE A. LISTE DES MÉTRIQUES UTILISÉES POUR LA CLASSIFICATION

MLC	L2SCA	Syntaxique	Longueur moyenne des propositions
C_S	L2SCA	Syntaxique	Ratio de propositions par phrases
CP_C	L2SCA	Syntaxique	Nombre de groupes coordonnés par proposition
swordtypes	LCA	Lexicale	Proportion de types de mots sophistiqués (à partir d'une liste de mots)
slextypes	LCA	Lexicale	Proportion de mots lexicaux (types) sophistiqués (à partir d'une liste de mots)
swordtokens	LCA	Lexicale	Proportion de mots(tokens) sophistiqués (à partir d'une liste de mots)
slextokens	LCA	Lexicale	Proportion de mots lexicaux (tokens) sophistiqués (à partir d'une liste de mots)
ld	LCA	Lexicale	Densité lexicale
ls1	LCA	Lexicale	Score de sophistication lexicale 1
ls2	LCA	Lexicale	Score de sophistication lexicale 2
vs1	LCA	Lexicale	Score de sophistication verbale 1
vs2	LCA	Lexicale	Score de sophistication verbale 2
cvs1	LCA	Lexicale	Score de sophistication verbale 3
ndwz	LCA	Lexicale	Variante du score NDW (Number of different words)
ttr	LCA	Lexicale	Type-to-token ratio

ANNEXE A. LISTE DES MÉTRIQUES UTILISÉES POUR LA CLASSIFICATION

msttr	LCA	Lexicale	Variation du type-to-token ratio
cttr	LCA	Lexicale	Variante du type-to-token ratio
rttr	LCA	Lexicale	Variante du type-to-token ratio
logttr	LCA	Lexicale	Variante du type-to-token ratio
uber	LCA	Lexicale	Score de diversité lexicale
lv	LCA	Lexicale	Score de variation lexicale
vv1	LCA	Lexicale	Score de variation verbale (types de verbes / nombre total de verbes)
svv1	LCA	Lexicale	Variante de vv1
cvv1	LCA	Lexicale	Variante de vv1
vv2	LCA	Lexicale	Score de variation verbale (types de verbes / nombre total de mots)
nv	LCA	Lexicale	Score de variation nominale
flesch_reading_ease	Textstat	Lisibilité	Score fondé sur la longueur des phrases et des mots
smog_index	Textstat	Lisibilité	Score fondé sur la longueur des phrases et la proportion de mots polysyllabiques
flesch_kincaid_grade	Textstat	Lisibilité	Score fondé sur la longueur des phrases et des mots
coleman_liau_index	Textstat	Lisibilité	Score fondé sur le nombre moyen de lettres et de phrases pour 100 mots
automated_readability_index	Textstat	Lisibilité	Score fondé sur le nombre de caractères par mot et de mots par phrase

ANNEXE A. LISTE DES MÉTRIQUES UTILISÉES POUR LA CLASSIFICATION

dale_chall_readability_score	Textstat	Lisibilité	Score fondé sur la proportion de mots "difficiles" (à partir d'une liste)
difficult_words	Textstat	Lisibilité	Score fondé sur la proportion de mots "difficiles" et la longueur des phrases (à partir d'une liste)
lin-sear_write_formula	Textstat	Lisibilité	Score fondé sur la proportion de mots de 2 syllabes ou moins ("faciles") et de mots de plus de 3 syllabes ("difficiles")
text_standard	Textstat	Lisibilité	Score fondé sur tous les scores ci-dessus

Métriques "microsystems" calculés sur les textes EN-PLAIN et EN-LEX

Le tableau suivant présente certaines des métriques du L2SCA modifié (*microsystem features*) calculées sur les textes en anglais¹.

FIGURE A.1

30	N_PREP_N	Functional	L2SCA_microsystem	Two noun linked by a preposition
31	N_N	Functional	L2SCA_microsystem	Compound nouns
32	MD_WILL	Functional	L2SCA_microsystem	Modal
33	MD_MAY	Functional	L2SCA_microsystem	Modal
34	MD_CAN	Functional	L2SCA_microsystem	Modal
35	MD_MUST	Functional	L2SCA_microsystem	Modal
36	MD_SHALL	Functional	L2SCA_microsystem	Modal
37	MD_WOULD	Functional	L2SCA_microsystem	Modal
38	MD_MIGHT	Functional	L2SCA_microsystem	Modal
39	MD_COULD	Functional	L2SCA_microsystem	Modal
40	MD_SHOULD	Functional	L2SCA_microsystem	Modal
41	MD_HAVETO	Functional	L2SCA_microsystem	Modal
42	MD_OUGHTTO	Functional	L2SCA_microsystem	Modal
43	ART_A	Functional	L2SCA_microsystem	Article
44	ART_THE	Functional	L2SCA_microsystem	Article

1. Une description complète des 102 métriques est disponible à l'adresse suivante : <https://www.iris-database.org/iris/app/home/detail?id=york:939367>

Annexe B

Matrices de confusion obtenues
suite à la la classification
automatique

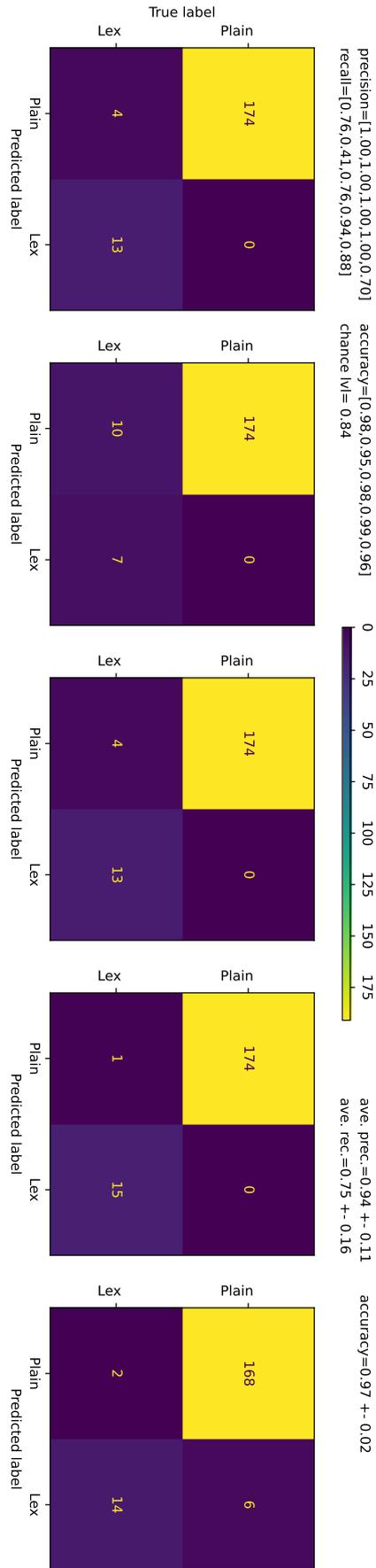
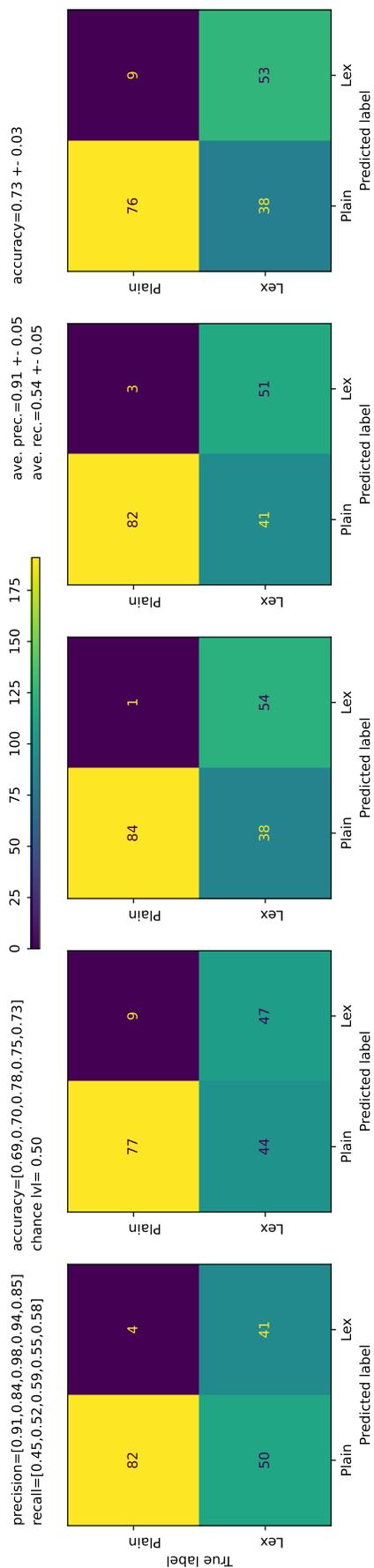


FIGURE B.1 – Matrice de confusion de la classification de EN-LEX et EN-PLAIN

FIGURE B.2 – Matrice de confusion de classification de FR-LEX et FR-PLAIN



Annexe C

Listes de mots utilisées pour l'analyse lexicale et terminologique

Liste de mots non-approuvés en anglais, adaptée de Plain English Campaign (2001)

absence	advance	apprise	circumvent	conjunction
abundance	advantageous	appropriate	clarification	consequently
accede	advise	approximately	combine	considerable
accelerate	affix	ascertain	combined	constitutes
accentuate	afforded	assemble	commence	construe
accommodation	aforesaid	assistance	communicate	consult
accompanying	aggregate	attempt	competent	consumption
accomplish	aligned	attend	compile	contemplate
accordance	alleviate	attributable	complete	contrary
accordingly	allocate	authorise	completion	correct
acknowledge	alternative	authority	comply	correspond
acquaint	alternatively	axiomatic	component	counter
acquiesce	ameliorate	behalf	comprises	courteous
acquire	amendment	belated	compulsory	cumulative
addition	anticipate	beneficial	conceal	currently
additional	apparent	bestow	concerning	customary
adjacent	applicant	breach	conclusion	deduct
adjustment	application	calculate	concur	deem
admissible	appreciable	cease	condition	defer

ANNEXE C. LISTE POUR L'ANALYSE LEXICALE ET TERMINOLOGIQUE

deficiency	encounter	furthermore	mandatory	peruse
delete	endeavour	generate	manner	place
demonstrate	enquire	grant	manufacture	pleasure
denote	enquiry	grounds	marginal	possess
depict	ensure	henceforth	material	possessions
designate	entitlement	hereby	materialise	practically
desire	envisage	herein	merchandise	predominant
despatch	equivalent	hereinafter	mislaid	prescribe
despite	erroneous	hereof	modification	preserve
determine	establish	hereto	moreover	previous
detrimental	evaluate	heretofore	negligible	principal
difficulties	evinced	hereunder	nevertheless	prior
diminish	ex	herewith	notify	proceed
disburse	exceptionally	hitherto	notwithstanding	procure
discharge	excessive	illustrate	number	profusion
disclose	exclude	immediately	numerous	prohibit
disconnect	excluding	implement	objective	projected
discontinue	exclusively	imply	obligatory	prolonged
discrete	exempt	inappropriate	obtain	promptly
discretion	expedite	inception	occasion	promulgate
discuss	expeditiously	incorporating	occasioned	proportion
dispatch	expenditure	incur	officio	provide
disseminate	expire	indicate	operate	provided
documentation	extant	inform	optimum	provisions
domiciled	extremity	initially	option	proximity
dominant	fabricate	initiate	ordinarily	purchase
drawn	facilitate	insert	otherwise	pursuant
due	factor	instances	outstanding	reconsider
duration	failure	intend	owing	reduce
dwelling	finalise	intimate	partially	reduction
economical	following	irrespective	participate	referred
eligible	formulate	jeopardise	particulars	refers
elucidate	forthwith	large	perform	regard
emphasise	forward	lieu	permissible	regarding
empower	frequently	locality	permit	regards
enable	furnish	locate	personnel	regulation
enclosed	further	magnitude	persons	reimburse

reiterate	residence	subject	therein	unoccupied
relating	restriction	submit	thereof	upon
remain	retain	subsequent	thereto	utilisation
remainder	review	subsequently	thus	utilise
remittance	revised	substantial	transfer	variation
remuneration	scrutinise	substantially	transmit	virtually
render	select	sufficient	ultimately	visualise
report	settle	supplement	unavailability	whatsoever
represents	similarly	supplementary	undernoted	whenever
request	solely	supplytenant	undersigned	whereas
require	specified	terminate	undertake	zone
requirements	state	thereafter	uniform	
reside	statutory	thereby	unilateral	

Liste des termes juridiques en anglais

Absentee	Attorney	Cohabit	Defence	Execution
Access	Bail	Commission	Defendant	Exhibit
Accused	Bankrupt	Competence	Dependant	Expert
Acquittal	Bankruptcy	Complainant	Deponent	Factum
Act	Bar	Conference	Deposition	Fees
Action	Barrister	Confession	Disclosure	Fiduciary
Adjournment	Bench	Consent	Discovery	Fine
Adoption	Beneficiary	Contest	Dismissal	Garnishee
Affidavit	Bill	Contract	Disposition	Garnishment
Affirmation	Binding	Conviction	Diversion	Guardian
Annulment	Bond	Costs	Divorce	Guilty
Answer	Brief	Counsel	Docket	Hearing
Appeal	CAS	Count	Duress	Homicide
Appellant	Case	Counterclaim	Election	Illegal
Apprehend	Caveat	Court	Endorse	Imprisonment
Arbitration	Certificate	Creditor	Endorsement	Incarceration
Arbitrator	Certiorari	Crown	Enforcement	Indictment
Arraignment	Charge	Custody	Estate	Injunction
Arrears	Charter	Damages	Estreatment	Inquest
Arrest	Child	Debtor	Evidence	Insolvency
Assessment	Claim	Declaration	Examination	Insolvent

ANNEXE C. LISTE POUR L'ANALYSE LEXICALE ET TERMINOLOGIQUE

Interpleader	Motion	Proceeding	Sealed	Testimony
Intestate	Neglect	Prohibition	Search	Tort
Judge	Negligence	Property	Security	Transcript
Judgment	Notary	Prosecute	Seizure	Trial
Jurisdiction	Notice	Prosecution	Sentence	Trust
Jury	Oath	Prosecutor	Separation	Trustee
Justice	Omission	Quash	Service	Uncontested
Law	Order	Questioning	Sheriff	Undertaking
Lawyer	Paralegal	Recipient	Solicitor	Unlawful
Leave	Pardon	Recognizance	Spouse	Vary
Legislation	Parole	Registrar	Standing	Verdict
Liability	Party	Relief	Statute	Waive
Litigation	Payor	Remand	Submission	Warrant
Mandamus	Perjury	Remedy	Subpoena	Withdrawal
Master	Plaintiff	Reply	Support	Witness
Mediation	Plea	Representative	Surcharge	Writ
Mediator	Pleading	Rescind	Tariff	
Minor	Precedent	Respondent	Testament	
Misrepresentation	Probate	Restitution	Testate	
Mitigate	Probation	Seal	Testator	

Liste des termes juridiques en français

absent	annuler	avocat	cohabiter	créancier
absolution	appel	barème	compétence	débiteur
absolution	appelant	barreau	conférence	décision
accusation	appréhender	bénéficiaire	confession	déclaration
accusé	arbitrage	bordereau	confiscation	dédommagement
acquittement	arbitre	bref	conjoint	défendeur
action	arrêter	capacité	consentement	défense
adolescent	arriéré	cause	contact	déjudiciarisation
adoption	assignation	caution	contester	délibéré
affidavit	atténuer	cautionnement	contrainte	demande
ajournement	audience	certificat	contrat	demandeur
aliments	auteur	certiorari	coupable	dénonciation
amende	autorisation	Charte	cour	dépens
annulation	avis	choix	Couronne	déposition

ANNEXE C. LISTE POUR L'ANALYSE LEXICALE ET TERMINOLOGIQUE

divorce	illicite	mémoire	poursuivant	retrait
divulgation	incarcération	mineur	poursuivre	SAE
doute	infraction	modification	précédent	saisie
droit	injonction	modifier	précisions	saisie-arrêt
emprisonnement	inobservation	motion	preuve	sceau
endosser	insolvabilité	négligence	probation	scellé
enfant	insolvable	notaire	procès	séparation
engagement	instance	novo	procuration	serment
enquête	interpellation	observations	procureur	shérif
évaluation	interpleader	omission	prohibition	signification
exécution	interrogatoire	ordonnance	promesse	succession
exécutoire	intestat	parajuriste	protonotaire	suppléant
expert	intimé	parental	puîné	suramende
facie	juge	parjure	rea	sûreté
failli	jugement	partie	recours	sursis
faillite	jurisprudence	partie	règlement	témoignage
fardeau	jury	passif	réhabilitation	témoin
fiducie	justice	payeur	rejet	testament
fouille	législation	peine	renoncer	testamentaire
garde	litige	perquisition	renvoyer	testateur
greffier	loi	pièce	réponse	transcription
habeas	magistrature	plaidoyer	représentant	verdict
homicide	mandamus	plaignant	requérant	warranto
honoraires	mandat	plan	requête	
huis	médiateur	possessoire	résilier	
illégal	médiation	poursuite	responsabilité	

Annexe D

Textes pour l'étude de cas sur les procès sémantiques

Nous incluons dans cette annexe des textes ou extraits des textes utilisés pour le codage des procès sémantiques. Le code-couleur est présenté ci-dessous.

Rose : procès relationnels

Vert : procès matériels

Bleu : Procès verbaux

Jaune : procès mentaux non verbaux

Gris : prédicats complexes

Souligné : auxiliaire modal

Gras : Passif

Procès sémantiques dans EN-CA-Jugements et FR-CA-Jugements

FIGURE D.1 – Procès sémantiques dans un extrait des motifs en anglais

Analysis

Section 15 of the Charter **states**:

Every individual **is** equal before and under the law and **has** the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination **based** on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability.

When **assessing** a claim under 15, this Court's jurisprudence **establishes** a two-step approach: Does the challenged law, on its face or in its impact, **draw a distinction based** on an enumerated or analogous ground, and, if so, does it **impose** "burdens or [deny] a benefit in a manner that **has** the effect of **reinforcing**, **perpetuating** or **exacerbating** disadvantage", **including** "historical" disadvantage? (Quebec (Attorney General) v A, [2013] 1 61, at paras. 323-24 and 327; Kahkewistahaw First Nation v Taypotat, [2015] 2 548, at paras. 19-20.)

FIGURE D.2 – Procès sémantiques dans un extrait des motifs en français

Analyse

Le paragraphe 15(1) de la Charte est **libellé** comme suit :

La loi ne **fait acception** de personne et **s'applique** également à tous, et tous **ont** droit à la même protection et au même bénéfice de la loi, indépendamment de toute discrimination, notamment des discriminations **fondées** sur la race, l'origine nationale ou ethnique, la couleur, la religion, le sexe, l'âge ou les déficiences mentales ou physiques.

Pour l'examen d'une demande **fondée** sur le par. 15(1), la jurisprudence de notre Cour **établit** une démarche en deux étapes : à première vue ou de par son effet, la loi contestée **établit-elle** une **distinction fondée** sur un motif énuméré ou analogue, et, dans l'affirmative, **impose-t-elle** « un fardeau ou **[nie-t-elle]** un avantage d'une manière qui **a** pour **effet** de **renforcer**, de **perpétuer** ou d'**accentuer** le désavantage», y compris le désavantage «historique» **subi**? (Québec (Procureur général) A, [2013] 1 S 61, par. 323324 et 327; Première Nation de Kahkewistahaw Taypotat, [2015] 2 S 548, par. 1920.)

La première question qui **se pose est** donc celle de **savoir** si la restriction contestée par les syndicats — en l'occurrence le délai de six ans **établi** à l'art 38 pour **mettre en œuvre** l'équité salariale dans le cas des femmes **occupant** un emploi dans des milieux sans comparateurs masculins — **établit** une **distinction fondée** sur le sexe.

Procès sémantiques dans les échantillons d'EN-CA-Résumés et FR-CA-Résumés

FIGURE D.3 – Procès sémantiques dans un extrait des résumés en anglais

A Quebec law **delaying** pay equity for women in female-dominated workplaces **violated** Charter equality rights, the Supreme Court has **ruled**, but it **declined** to **strike** the law down.

In 1996, Quebec **passed** a law that **required** all employers with ten or more employees to **ensure** pay equity for women by 2001. Pay equity **means** equal pay for work of equal value. For most employers, the 1996 law **meant** **identifying** jobs that are **done** mostly by women, and **comparing** their salaries to the salaries for jobs **done** mostly by men in the same workplace. The problem **was** that there **was** no way to **determine** pay equity for women in workplaces where there **were** no mostly male jobs for comparison.

Measuring pay equity in those workplaces **was** difficult. Quebec **gave** its Pay Equity Commission more time to **come up** with a solution. This **meant** that women in mostly-female workplaces **waited** six years longer for pay equity than women in mixed-gender workplaces. Some unions **went** to court to **complain** that this delay **discriminated** against the women who were **forced** to wait. The trial judge **found** that the delay did not **violate** the women's equality rights. This **was** because the **distinction** the government **made** was not **based** on the women's sex (a ground of discrimination **listed** in the Canadian Charter of Rights and Freedoms). Instead, it **was** because there **was** no male group to **compare** them to. The Quebec Court of Appeal **agreed**.

FIGURE D.4 – Procès sémantiques dans un extrait des résumés en français

La Cour suprême **statue** qu'une loi québécoise **différant** l'accès à l'équité salariale des femmes **travaillant** dans des milieux à prédominance féminine **porte atteinte** aux droits à l'égalité **garantis** par la Charte, mais **refuse** de l'invalider.

En 1996, le Québec a **adopté** une loi qui **obligeait** tous les employeurs comptant dix salariés ou plus à faire en sorte que les femmes **obtiennent** l'équité salariale au plus tard en 2001. L'équité salariale **s'entend** d'un salaire égal pour un travail de valeur égale. La plupart des employeurs **devaient repérer** les emplois **occupés** principalement par des femmes et **comparer** les salaires correspondants à ceux **versés** pour des emplois **occupés** principalement par des hommes dans le même lieu de travail. Un problème **demeurait** toutefois : on ne **pouvait déterminer** les ajustements salariaux requis dans le cas des femmes qui **travaillaient** dans des secteurs **dépourvus** d'emplois à prédominance masculine permettant la comparaison.

Évaluer l'équité salariale dans ces milieux de travail **était** difficile. Le Québec a donc **accordé** à sa Commission de l'équité salariale un délai pour **trouver** une solution. Les femmes qui **travaillaient** dans des secteurs à prédominance féminine ont donc **attendu** six ans de plus que celles qui **travaillaient** dans des secteurs à plus grande mixité pour **obtenir** l'équité salariale. Des syndicats ont **fait valoir** devant les tribunaux que le délai **était** discriminatoire vis-à-vis des femmes en cause.

Procès sémantiques dans l'échantillon d'EN-Lois et FR-Lois

FIGURE D.5 – Procès sémantiques dans un extrait de l'Equality Act 2010

CHAPTER 2
PROHIBITED CONDUCT
Discrimination

13 Direct discrimination

A person A **discriminates** against another B if, because of a protected characteristic, A **treats** B **less favourably** than A **treats** or would **treat** others.

If the protected characteristic **is** age, A does not **discriminate** against B if A **can show** A's treatment of B to **be** a proportionate means of achieving a legitimate aim.

If the protected characteristic **is** disability, and B **is** not a disabled person, A **does not discriminate** against B only because A **treats** or would **treat** disabled persons more favourably than A treats B.

If the protected characteristic **is** marriage and civil partnership, this section **applies** to a contravention of B who **is** married or a civil partner.

If the protected characteristic **is** race, less favourable treatment **includes** **segregating** B from others.

If the protected characteristic **is** sex-
less favourable treatment of a woman **includes** less favourable treatment of her because she is **breast-feeding**;
in a case where B **is** a man, **no account is to be taken** of special treatment **afforded** to a woman in connection with pregnancy or childbirth.

FIGURE D.6 – Procès sémantiques dans un extrait du Code Pénal français

Constitue une discrimination toute **distinction opérée** entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de la particulière vulnérabilité **résultant** de leur situation économique, apparente ou **connue** de son auteur, de leur patronyme, de leur lieu de résidence, de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur capacité à s'**exprimer** dans une langue autre que le français, de leur appartenance ou de leur non-appartenance, vraie ou **supposée**, à une ethnie, une Nation, une prétendue race ou une religion déterminée.

Constitue également une discrimination toute distinction **opérée** entre les personnes morales sur le fondement de l'origine, du sexe, de la situation de famille, de la grossesse, de l'apparence physique, de la particulière vulnérabilité **résultant** de la situation économique, apparente ou **connue** de son auteur, du patronyme, du lieu de résidence, de l'état de santé, de la perte d'autonomie, du handicap, des caractéristiques génétiques, des mœurs, de l'orientation sexuelle, de l'identité de genre, de l'âge, des opinions politiques, des activités syndicales, de la capacité à s'**exprimer** dans une langue autre que le français, de l'appartenance ou de la non-appartenance, vraie ou **supposée**, à une ethnie, une Nation, une prétendue race ou une religion déterminée des membres ou de certains membres de ces personnes morales.

Procès sémantiques dans l'échantillon d'EN-CAB et FR-DILA

FIGURE D.7 – Procès sémantiques dans un extrait d'EN-CAB sur la discrimination

Direct discrimination

If you've been **treated** unfairly by someone simply because of who you **are**, this **could be** direct discrimination. Direct discrimination **is** against the Equality Act 2010. Discrimination which **is** against the Equality Act **is** unlawful. This **means** you can take action in the civil courts. If you've **experienced** unlawful discrimination, you **may be** able to **do** something about it. **Read** this page to **find out** more about direct discrimination.

Remember direct discrimination can **be** because of:

who you **are**

who someone **thinks** you **are**

someone you're with.

It's only unlawful discrimination if you're **treated** differently and worse because of a protected characteristic.

What **is** direct discrimination?

Direct discrimination **is** when you're **treated** differently and worse than someone else for certain reasons. The Equality Act **says** you've been **treated** less favourably.

FIGURE D.8 – Procès sémantiques dans un extrait de FR-DILA sur la discrimination

Discrimination à la location (logement privé)

Le propriétaire ne **peut pas refuser** un locataire en raison de son origine ethnique, sa nationalité ou son orientation sexuelle. C'est un cas de discrimination **puni** par la loi.

Critères de sélection autorisés

Pour sélectionner un locataire parmi les candidats, le propriétaire a le droit de **se fonder** sur des critères objectifs **ayant trait** à sa situation financière comme :

Le revenu

La présence ou non d'une personne se pourtant caution,

La situation professionnelle (CDI, CDD) etc ...

Attention :

le propriétaire n'a pas le droit de **réclamer** certains documents privés comme un relevé de compte.

Cas de discrimination

Certains critères ne **doivent pas être retenus** pour choisir un locataire, car ils **sont** discriminatoires :

Origine géographique, nom de famille, lieu de résidence

Appartenance ou non-appartenance, réelle ou supposée, à une ethnie ou à une nation

Langue parlée (autre que le français),

Index

A

apprentissage supervisé, 159

C

classification automatique, 158

collocations, 64

common law, 35

complexité, 108

 complexité lexicale, 116

 complexité syntaxique, 142

D

discours du droit, 41

 discours administratif, 48

 discours judiciaire, 45

 discours législatif, 43

G

genre discursif, 37

L

lexique juridique, 54

M

médiation juridique, 70

métriques de complexité, 138

P

passif, 123

phraséologie, 64

procès sémantiques, 135

prédicat complexe, 62

S

schéma lexico-grammatical, 65

scores de lisibilité, 139

style clair, 28

T

terminologie juridique, 52