



**HAL**  
open science

# The Phraseology of Legal French and Legal Popularisation in France and Canada: A Corpus-Assisted Analysis

Manon Bouyé, Christopher Gledhill

► **To cite this version:**

Manon Bouyé, Christopher Gledhill. The Phraseology of Legal French and Legal Popularisation in France and Canada: A Corpus-Assisted Analysis. *Languages*, 2024, 9 (3), pp.107. 10.3390/languages9030107. hal-04595282

**HAL Id: hal-04595282**

**<https://u-paris.hal.science/hal-04595282>**

Submitted on 31 May 2024

**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.

## Article

# The Phraseology of Legal French and Legal Popularisation in France and Canada: A Corpus-Assisted Analysis

Manon Bouyé <sup>1,†</sup>  and Christopher Gledhill <sup>2,\*,†</sup> 

<sup>1</sup> Equipe LEADS, Département d'Enseignement et de Recherche Langues, École Normale Supérieure Paris-Saclay, 91190 Gif-sur-Yvette, France; manon.bouye@ens-paris-saclay.fr

<sup>2</sup> Laboratoire CLILLAC-ARP, UFR EILA, Faculté Sociétés et Humanités, Université Paris Cité, 75006 Paris, France

\* Correspondence: christopher.gledhill@u-paris.fr

† These authors contributed equally to this work.

**Abstract:** The popularisation of legal knowledge is a critical issue for equal access to law and justice. Legal discourse has been justly criticised for its obscure terminology and convoluted phrasing, which notably led to the Plain Language Movement in English-speaking countries. In Canada, the concept of Plain Language has been applied to French since the 1980s due to the official policy of bilingualism, while the concept has only been recently discussed in France. In this paper, we examine the impact of Plain Language rewriting on legal phraseology in French popularisation contexts. The first aim of our study is to see if plain texts published in France contain more traces of legal phraseology than French Canadian texts. Our second objective is to determine if a 'phraseology of plain language' can be identified across genres and languages. To do this, we compare two corpora of expert-to-expert legal texts written in French—made up, respectively, of legislative texts published in France and judicial texts published by the Supreme Court of Canada—with two corpora of texts that are claimed to have been written in Plain French Language for a non-expert readership—texts that guide laypersons through legal and administrative processes in France and summaries of decisions by the Supreme Court of Canada. Using n-grams, we extract and discuss the patterns that emerge from the corpora. In particular, our analyses rely on the concept of 'lexico-grammatical patterns', defined as the minimal unit of meaningful text made up of recurrent sequences of lexical and grammatical items. We then identify a sample of recurring lexico-grammatical patterns and their discursive functions.

**Keywords:** legal French; plain French; plain language; popularisation; phraseology



**Citation:** Bouyé, Manon, and Christopher Gledhill. 2024. The Phraseology of Legal French and Legal Popularisation in France and Canada: A Corpus-Assisted Analysis. *Languages* 9: 107. <https://doi.org/10.3390/languages9030107>

Academic Editors: Jeanine Treffers-Daller, Julien Longhi and Nadia Makouar

Received: 15 November 2023

Revised: 8 February 2024

Accepted: 26 February 2024

Published: 19 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Plain Language (PL) is an attempt to encourage official institutions and other organisations to communicate with laypeople using accessible, clear, user-friendly language. The PL movement first gained momentum in the 1970s and 1980s in English-speaking countries: notably, in the United Kingdom, Australia, and New Zealand and then the United States and Canada (Asprey 2004). The initial domain of application for PL was legal and judicial contexts, such as the drafting of contracts and statutes, but the concept has since spread to other areas, such as public administration and medicine. PL differs from more formal language schemes (such as Basic English, Controlled Language, and so on) in that it corresponds to a nebulous series of stylistic preferences rather than an explicitly defined set of re-writing rules or vocabulary. The guidelines for PL include negative advice (avoid the passive, do not use rare or specialised terms, avoid complex verbs and complex prepositions, etc.) as well as more positive recommendations (use shorter sentences, prefer direct expressions, address the reader as 'you', etc.) (Cutts 2008; Williams 2004). Notwithstanding a lack of formal definition, PL has attained a high degree of official recognition in various English-speaking countries and has been implemented in both expert-to-expert communication (as in statutes) (Williams 2004, 2015) and expert-to-non-expert communication.

Many studies have been devoted to discussing the implementation of PL (Williams 2015, 2022) or its reception in various legal settings in English-speaking countries, including the United Kingdom and New Zealand (Masson and Waldron 1994; Rossetti et al. 2020). There has also been research from the point of view of discourse analysis that looks at popularisation and the dissemination of legal knowledge. In this paper, we focus on the popularisation of legal knowledge, defined by Engberg et al. (2018) as the recontextualisation of legal knowledge from contexts that exhibit a form of power asymmetry to a new non-expert context, with the intent of adapting the presentation of knowledge to the audience (Engberg et al. (2018) [our paraphrase]). As can be seen in this volume, numerous authors have investigated popularisation of the law carried out by legal institutions in various languages such as English (Cacchiani et al. 2018; Turnbull 2018), German (Luttermann, and Engberg 2018) and French (Preite 2016, 2018), while others have examined popularisation in non-institutional contexts through various media, including studies of YouTube videos used by expert lawyers (Cavaliere et al. 2018), children's books (Diani 2018) or teaching applications of TV shows in legal English classes (Dąbrowski 2017). Our paper follows on from research on popularisation produced by legal institutions, as it focuses on legal information texts published by the French government and Plain Language Summaries of judgements published by the Supreme Court of Canada. Both genres are intended for a non-specialised audience, and to the best of our knowledge, these genres have seldom been compared.

Generally speaking, the concept of PL is less well established outside the English-speaking world, and there have consequently been far fewer studies on PL or on the principles of clear language, especially in languages such as French. However, it is notable that the principles of PL have been widely adopted and implemented in Canada, and therefore also in the French-speaking parts of Canada, as part of bilingual language policy under the name *Langue Claire et Simple* (Simple and Clear Language). Various legal professions (barristers) and institutions (the Supreme Court of Canada) now claim to use PL in Canada (Asprey 2004). In contrast, the concept of PL is not as well-developed in France, and it has not achieved the same level of institutional recognition. One reason for this may be that relations between French citizens and the administration (*Service Public*) are notoriously difficult. This has been argued by both linguists (Collette et al. 2002) and independent observers, who have pointed out the complexity of legal and administrative procedures for French users of the law (des Droits 2019). Thus, although there have been attempts to implement the principles of PL in some contexts, it strikes us that there is generally still a considerable gap between the user-friendly discourse adopted by 'public-facing' organisations in many English-speaking contexts and the highly elaborate 'techno-heavy' style of French official discourse.

These general observations lead us to test two assumptions in this paper. In the first instance, we set out to test the hypothesis that French texts from France (FR FRA) are 'less simplified' than comparable texts from English-speaking countries (EN UK, EN NZ, etc.). This hypothesis was tested and partly confirmed in Bouyé (2022). In this paper, we test the related hypothesis that Canadian-French texts (FR CAN) are 'more simplified' than their European French counterparts (FR FRA). By 'more or less simplified', we are not talking about a single quantifiable characteristic, but rather, we are talking about two different configurations that can be identified systematically in two types of discourse (expert legal texts vs. plain legal texts). More specifically, abundant research has established that legal language is characterised by a 'highly nominal' style (Crystal and Davy 1969)—one of the characteristics that legal French shares with English, amongst other languages—along with Latin and Latinate terms, set formulae, a formal register, complex syntax due to a high degree of subordination, and the use of complex prepositional phrases (Galonnier 1997). One of our objectives is to examine the impact of the reconfiguration of legal knowledge on certain syntactic and stylistic features, including nominal and prepositional forms, as PL guidelines often encourage the use of verbal rather than nominal forms (Plain English Campaign 2022).

In the remaining sections of this paper, we explore these questions quantitatively and through the prism of phraseology. Several analysts have examined the phraseology of legal language, with a number of studies looking at regular expressions and lexical bundles in English (Biel 2017; Breeze 2017; Goźdź Roszkowski and Pontrandolfo 2013; Goźdź-Roszkowski and Pontrandolfo 2017). However, fewer studies have been conducted on the phraseology of PL in French. What we mean by ‘the phraseology of plain language’ is simply the typical wordings (routine formulae, extended collocations, lexico-grammatical patterns, etc.) that can be seen as statistically significant in one type of text (popularized texts, mediated knowledge, etc.) when compared to other types of text. More specifically, our concept of ‘phraseology’ corresponds to recurring sequences of lexico-grammatical sequences that operate as whole semantic units and serve a regular discourse function within a specific type of discourse or genre, a notion we have explored elsewhere (Bouyé and Gledhill 2019). Thus, in popularised legal texts, it is possible to identify recurrent sequences of this type (*This is called . . . Dans un délai de*) and to associate these sequences with specific discourse functions. These discourse functions include referential functions, i.e., sequences that refer to participants or elements of the legal process, as well as metatextual functions, such as sequences that define a term or direct the reader towards another part of the text or towards another text. In one of the first studies of this type (Bouyé and Gledhill 2019), we attempted to set out some characteristics of PL phraseology in English and French using n-grams. In this paper, we return to the concept of the ‘lexico-grammatical pattern’ (LxGr) in order to examine whether there is such a phenomenon as the ‘phraseology of simplification’. In order to grasp the implications of this, it is important to provide a more formal account of LxGr patterns. We define LxGr patterns (Gledhill et al. 2017) as recurrent sequences of lexical items (‘collocations’) that correspond to regular grammatical structures<sup>1</sup> and that have a recognisable frame of reference or discourse function. Thus, each LxGr pattern corresponds to a ‘minimal meaningful unit of text’. Unlike n-grams and other fixed sequences, LxGr patterns are productive and potentially discontinuous. The simplest forms of LxGr patterns are routine formulae or ‘speech acts’ (such as greetings, warnings, official pronouncements, etc.) (Gledhill et al. 2017).

In addition to our hypotheses regarding Plain Language in legal discourse, in this paper we also examine a number of more general research questions regarding LxGr patterns. In particular: what is the smallest possible sequence of items (n-gram) that corresponds to an LxGr pattern (i.e., a meaningful unit of text)? Furthermore, given a random selection of n-grams, is it possible to predict the discourse function for that LxGr (e.g., definition, procedure, explanation, evaluation, etc.)?

Returning more specifically to the topic of Plain Language in French, this paper considers the following research questions:

- Are there characteristic LxGr patterns in legal texts (i.e., in non-plain legal texts)?
- Are there traces of such patterns in PL texts?
- Similarly, are there characteristic LxGr patterns of PL in administrative discourse?
- More specifically, is it possible to establish a difference between generic phraseology (belonging to several ‘genres’) and specific phraseology (patterns that are ‘unique’ or at least more salient in one genre as opposed to all the others)?

We propose to answer these questions in the following sections. Section 2 introduces our data and corpus tools. Sections 3 and 4 present, analyse and discuss the results obtained from our data.

## 2. Materials and Methods

### 2.1. Data

The textual data used in this study are based on two French-language corpora: one consisting of popularisation texts destined for non-expert law users, entitled PLAIN, and the second, entitled LEX, made up of expert-to-expert written legal genres. Each of these corpora is subdivided into two subcorpora.

Concerning the PLAIN corpus, as mentioned above, this paper focuses on the popularisation of legal knowledge in French published by two government institutions. [Turnbull \(2018\)](#) distinguishes between ‘popularisation’, defined as the recontextualisation of information with the aim of broadening the reader’s general knowledge, and ‘knowledge mediation’, in which information is transferred with the aim of allowing readers to take action performatively and thus to ‘empower’ themselves. The two popularisation genres represented here can be said to be instances of both mediation and popularisation as they relate directly to citizens’ access to justice (*accès au droit et à la justice*) and public understanding of rights (*connaître et faire valoir ses droits*) and aim to help their readers make sense of the decisions taken by major judicial institutions or guide them through various legal and administrative processes. The first PLAIN subcorpus is composed of texts published by one of the most popular public service dissemination websites in France: *Service Public*. The texts are drafted and published by the governmental agency *Direction de l’information légale et administrative* (DILA), which is a department of the French Prime Minister’s Office. It is one of the main public service legal mediators in the country. This subcorpus, entitled FR-Admin-DILA, is made up of 337 texts published between 2017 and 2019 and 466,472 word tokens. The second PLAIN subcorpus was collected from the Canadian Supreme Court website. It is made up of 66 summaries of decisions taken by the Canadian Supreme Court between 2017 and 2019. The small size of this corpus, which comprises 68,025 word tokens, can be explained by the specific type of text it is made of. The summaries, or ‘Cases in Brief’, are short summaries of the Court’s judgements that recall the facts, explain the final decision reached by the Court and explain the positions of both the majority and dissenting opinions. This corpus is called FR-CA-Résumés.

As for the LEX corpus, it comprises FR-LAW, which contains articles and excerpts from statutes drafted between 1967 and 2018 and which were still in force at the time the corpus was compiled. It is representative of the legislative register and contains 5,083,750 word tokens. The second specialised corpus, entitled CA-Judgements, is made up of judicial texts: namely, decisions written and delivered by the Justices from the Supreme Court of Canada published between 2017 and 2018. It contains a total of 682,294 word tokens.

## 2.2. Methodology

Although in this paper we are focusing on phraseology, the first step in our analysis is to establish candidate sequences in the form of n-grams. In this case, we are interested in n-grams that identify parts of speech (POS) rather than just lexical forms. To identify salient POS-grams, we use the n-gram function of the concordance software Sketch Engine. Part-of-speech tags (not words) are used as attributes in order to extract not only major lexical bundles but also possibly salient syntactic regularities that could characterise our corpora. POS-grams allow us to capture the most salient grammatical constructions that may be candidate forms for more ‘recognisable’ LxGr patterns (as mentioned below, not all n-grams are potential LxGr patterns). To extract the most salient POS-grams in these corpora, a general reference corpus was used. This was the French Web Corpus (Sketch Engine) 2017, which is part of the TenTen family, a set of corpora obtained through web-crawling ([Jakubíček et al. 2013](#)) that contains a variety of text types, including news texts. We consider the French TenTen to be a general reference corpus. It must be noted that although this reference corpus contains more than four billion word tokens, only the first million word tokens in the corpus are used when computing the frequency on Sketch Engine. Key POS-gram candidates were identified based on Sketch Engine’s keyness score feature, which uses the ‘simple math method’ ([Kilgarriff 2009](#)) to identify key words or key n-grams based on the normalised frequency of the word or n-gram in the focus corpus in relation to its normalised frequency in the reference corpus and includes a smoothing parameter. For each subcorpus, POS-grams with a keyness score above 100 were marked as LxGr-candidates. This means they were found to be at least a hundred times more frequent in the subcorpora than in the French TenTen reference corpus. To ensure that the selected patterns were over-represented in the legal or plain legal corpora as compared to

the reference corpus, chi-square tests were performed for each dataset using R. This point also applies to the other results we mention in Section 4 of this paper.

The n-gram function with POS tags as attributes returned a list of POS tag sequences that had to be converted back into readable POS patterns or lexical bundles. The analysis of concordances for each candidate POS-gram thus had to be carried out to identify LxGr patterns and their functions. Many POS-gram sequences with very high keyness scores corresponded to noise, i.e., they referred to numbers, prices or abbreviations in the various corpora. Others corresponded to parts of larger LxGr patterns. It was therefore necessary to consider POS-grams using contextual analysis and concordances. In Section 3, we present some POS-grams and patterns that were both highly recurring and interesting in terms of their rhetorical functions. This explains why some of our corpora have more distinctive patterns than others. As mentioned above, we do not attempt to make a simple distinction between ‘specialised’ and ‘popularised’ phraseology, but we are attempting to identify the typical patterns that emerge in comparable corpora, which can be characterised as expert-oriented (‘specialised’) and non-expert oriented (‘popularised’). In the following discussion, we see examples of salient phraseology that can be identified as typical (in a statistical sense) in one corpus and atypical (or absent) in the other or can be observed as occurring in both corpora. This does not mean that we are in a position to fully characterise the phraseology of simplified language; rather, we claim here simply that it is possible to identify a representative sample of the most salient (outstanding, archetypical) phraseological units in our corpora, thus paving the way for a more complete analysis using other methods (e.g., textometrics).

We then performed a qualitative analysis of salient sequences based on domain-specific discourse functions put forward by Goźdź-Roszkowski et al. (2012). These include the category ‘legal reference bundles’: in particular, Institutional bundles, which refer to institutions; or Terminological bundles, which refer to specialised terms. The second category that was used to classify the POS-grams is text-oriented bundles: in particular, Structuring bundles. The final category used by Goźdź-Roszkowski et al. (2012) is Stance-oriented bundles, which contains Attitudinal bundles and Epistemic Stance bundles. Not all of these categories are represented in the results below.

In the rest of this paper, we use the following typographic conventions to refer to LxGr patterns. Lexical items are shown in italics, and each LxGr pattern is presented between angular brackets < >. All frequencies given in Tables are relative frequencies per million words (pmw).

### 3. Results

#### 3.1. Analysis of Lexicogrammatical Patterns in Legal Texts

Table 1 shows some of the key POS-grams that were extracted from the FR-Law corpus from France.

**Table 1.** Key POS-grams in FR-Law Corpus (statutes).

Pattern	Relative Frequency in FR-Law Corpus	Relative Frequency in General Reference Corpus	Keyness Score
<Prep + <i>l'article</i> + Capital letter + Number <sup>1</sup> >	3151.6	4.8	540.8
<Prep + N masc + Prep + Det + N fem>	1545.1	64.3	23.6
<N + <i>prévues/fixées</i> + Prep + article + N>	795.3	2.5	231.8
< <i>le cas échéant</i> > <sup>2</sup>	623.36	1.76	

<sup>1</sup> Translation: ‘Prep + section/article + capital letter + number’. <sup>2</sup> Translation: ‘if/when applicable’.

Table 2 presents the key POS-grams in the corpus of Supreme Court judgements written in French.

**Table 2.** Key POS-grams in the CA-Judgements Corpus (Supreme Court judgements).

Pattern	Relative Frequency in CA-Judgements	Relative Frequency in General Reference Corpus	Keyness Score
<La Cour d'appel <sup>1</sup> >	648.2	0.4	461.5
<Prep + l'article + Num> <sup>2</sup>	965.3	0.011	955.6
<Les juges majoritaires> <sup>3</sup>	401.5	0.02	402.5 <sup>3</sup>
<Prep + det + N + prep + N>	612.6	3.7	129.9

<sup>1</sup> Translation: 'Prep + section/article + number'. <sup>2</sup> Translation: 'the Court of Appeal'. <sup>3</sup> Translation: 'the majority'.

Some patterns found in the results are specific (unique) to the type of discourse, or register, as can be seen when comparing the relative frequencies of these POS-grams in the two focus corpora, where they occur several hundred or even thousand times per million words, whereas they only appear a few times or a few dozen times in the general reference corpus. For example, Table 2 shows two specific 3-grams from the CA-Judgements corpus: *la Cour d'Appel*, which means 'the Court of Appeal' and *les juges majoritaires*, 'the majority judges'. Both are fragments of phrases that belong to longer Institutional bundles. These phrases refer to central figures and institutions of the Common law system: namely, the judges from the lower court along with the Court of Appeal, whose decision is the basis for a case being brought before the Supreme Court. In the case of the FR-Law corpus, one recurring pattern is *le cas échéant*, which means 'if the [aforementioned] conditions are fulfilled'. As we will see, these are key terms and phrases in legal discourse that can also be found in PLAIN texts but with only marginal recurrence (20 pmw).

What can be noted from Tables 1 and 2, however, is that both genres share common LxGr patterns, although with some variation. Most of these patterns correspond to what Goźdz-Roszkowski et al. (2012) calls Textual bundles and refer to other statutes or decisions or to other articles or sections in the same statute or judgement. These patterns are linked to cross-referencing: a broader characteristic of legal discourse (Tiersma 2000) that can be found in both genres (statute or Supreme Court opinion). In statutes, Bhatia (1994) calls these Referential provisions and explains that they point intertextually to other legislative texts or passages within the same text.

- (1) <Selon l'art. 662 > du Code criminel, la personne inculpée d'une infraction qui n'a été prouvée que partiellement peut être déclarée coupable d'une infraction moindre et incluse. (CA-Judgements)

**Section 662** of the Criminal Code provides that where a person is charged with one offence, but only a part of that offence is proved, he or she may be convicted of a lesser, included offence. (Official Supreme Court of Canada translation)

- (2) La durée de la période prévue <à l'article L. 434-9 > est fixée à trois ans. (FR-Law)

The duration of the period is set to three years **under art. L. 434-9**. (Our translation)

Such patterns (using a prepositional phrase in French and a thematised phrase in English) are pervasive in this type of expert discourse (statutes, decisions), which requires constant references to previous decisions, statutes and legal instruments.

Another particularly interesting candidate is represented by the sequence <Prep + N + prep + N + prep>. This is a good LxGr candidate since it consists of a highly regular recurring grammatical structure but also allows for a large amount of lexical variation. A quick examination of some concordance results reveals that this pattern corresponds to a complex prepositional sequence, i.e., a prepositional phrase that introduces another noun phrase or prepositional phrase(s).

- (3) Les présents pourvois portent <sur l'étendue de l'obligation de communication du ministère public en ce qui a trait aux> registres d'entretien des alcootests. (CA-Judgements)

These appeals deal **with the scope of the Crown's disclosure obligations with respect to** maintenance records of breathalyzer instruments. (Official translation by the Supreme Court of Canada)

The fact that this pattern is highly salient in our French legal corpora is significant. The use of complex noun chains and prepositional cascades is considered a typical feature of legal language, especially when it is in the form of complex prepositions in expressions such as *for the purpose of*<sup>2</sup> (Bhatia 1983; Biel 2017; Coulthard et al. 2016). Even more significantly, according to PL drafting guides in English and French, complex nominals and prepositional chains are among those features that should be avoided by legal writers because they purportedly contribute to the heavy nominal style (Crystal and Davy 1969) of legal language. In addition, the packing of information in nominal or prepositional cascades increases lexical density in texts (Halliday 1994). We discuss this pattern further in Section 3.3 since it is also found to some extent in PLAIN texts from our corpora.

It is also interesting to note at this point that many of the patterns we have identified above can be found within other patterns (either embedded or adjacent), as is evident in the example below, an excerpt from the French-Law corpus, in which two patterns from Table 1 can be identified:

- (4) *Tout travailleur de nuit bénéficie d'un suivi individuel régulier <de son état de santé> <dans les conditions fixées à l'article L. 4624-1>*. (French-Law corpus) Any night worker has the right to regular health check-ups pursuant to the conditions set out (Our translation)

The fact that patterns occur alongside or within other patterns (thus, we use the term 'chains of pattern' or cascades) corroborates the idea that phraseological patterns constitute the building block of specialised discourse and here, in particular, of legislative or judicial discourse. We now turn to the main LxGr patterns in the PLAIN corpus.

### 3.2. Analysis of Lexicogrammatical Patterns in PLAIN French Texts

#### 3.2.1. In the European French Admin Corpus

The key POS-grams in the French administrative corpus addressing law users correspond to very specific n-grams (or lexical bundles). Table 3 shows the most frequent patterns.

Table 3. Key LxGr patterns in French administrative texts (FR-PL-Admin).

Pattern	Frequency (pmw)	Freq in General French Corpus	Keyness Score
<Vous devez/pouvez <sup>1</sup> + V>	4032.4	14.3	263.3
<dans un délai de <sup>2</sup> + Numeral Adj + jours/mois/semaines>	568.1	0.9	297.7
<Prep + N + prep + N + prep>	508.1	3.6	111.3
<Où s'adresser ?> <sup>3</sup>	428.7	0.02	418.7
<de quoi s'agit-il ?> <sup>4</sup>	237.9	0.02	233.3

<sup>1</sup> Translation: 'You must/can'. <sup>2</sup> Translation: 'Within X days/months/weeks'. <sup>3</sup> Translation: 'Where can I get help?' <sup>4</sup> Translation: 'What does this mean?'

Although many of these items are very short, we suggest that several of these sequences are fragments of longer LxGr patterns, which themselves constitute recognisably meaningful units of text (according to our definition above). Many of these sequences correspond to text-oriented patterns, framed as direct questions, which have a cohesive, organisational function. These LxGr patterns either introduce the definition of a just-mentioned term, as in Examples 5 and 6 below, or explain who or what institution the law users can contact next in the legal process they are involved in. In Example 7, the question 'Who should I contact?' is immediately followed by a sentence that gives the phone number of the Police.

- (5) **<De quoi s'agit-il ?>** Le sursis simple dispense la personne condamnée de l'exécution de la peine prononcée (peine de prison et/ou d'amende). (FR-Admin-PL)  
**What does this mean?** Suspension with probation allows convicted individuals to avoid a particular sentence (imprisonment or fine) (Our translation).



- (6) <De quoi s'agit-il ?> Le contrôle judiciaire est une mesure qui soumet la personne mise en cause dans une affaire pénale à une ou plusieurs obligations, dans l'attente de son procès. (FR-Admin-PL)  
 <What does this mean?> Bail is a measure that can be given, with one of more conditions, to a person accused of a criminal offence while they wait for trial. (Our translation)
- (7) <Où s'adresser ?> Police secours – 17 Par téléphone Composez le 17 en cas d'urgence concernant un accident de la route, un trouble à l'ordre public ou une infraction pénale.  
 Where to get help Police services – By phone – Call 17 in case of an emergency, traffic accident, disturbing the peace or criminal offence. (Our translation) (FR-Admin-PL)

Taken as a whole, these LxGr patterns are part of what discourse analysts such as [Turnbull \(2018\)](#) have called 'conversational turns'. As such, they represent significant structuring devices in the dissemination of legal knowledge. Drafters of legal knowledge mediation texts imagine and anticipate the questions law users might have and use interrogative sentences to structure their texts when explaining the steps of the legal procedure or situations related to a legal right. In the case of *De quoi s'agit-il ?* (Example 5), we have an explicit signal that the text is going to provide the definition of a term. As we see later on, this structure is comparable to <il y a + N>, in which the authors interrupt their exposition to provide an explicit metadiscoursal definition. The difference here is that <De quoi s'agit-il?> and <Où s'adresser> both belong to oral discourse and are explicit markers of turn-taking; whereas <Il y a X quand / lorsque> belongs to elaborate, expository discourse.

Other patterns correspond to lexical phrases that are linked to the steps of the legal or administrative process itself, such as the complex prepositional phrase *dans un délai de + N*, which defines the time limit for legal action for law users. This highly frequent bundle, which appears only 0.9 pmw in the general French corpus, appears to be quite specific to the FR-Admin corpus. It is usually part of an extended LxGr pattern that contains an Actor (usually second-person *You*), a verb phrase referring to some type of legal action (such as reporting a crime or referring to an institution) and the complex prepositional phrase, which functions as a time adjunct.

- (8) *À savoir : en raison des règles de prescription, vous devez déposer votre plainte pour viol <dans un délai de 20 ans> à compter de la date des faits.* (FR-PL-Admin)  
 Please be aware that because of the statute of limitation, you need to file a rape complaint **within twenty years**. (Our translation)
- (9) *Si votre demande est acceptée, vous en êtes informé par courrier <dans un délai de 4 mois>.* (FR-PL-Admin)  
 If your request is approved, you will be notified **within four months**.

In Example 8, the time limit that is defined is twenty years, and the legal action is reporting a specific felony (rape). Interestingly, this type of pattern can also be found in the FR-Law corpus, as in the example below. Although the pattern <dans un délai de> appears in the legislative corpus, it is about half as frequent in the specialised corpus (229.5 pmw vs. 568.1 pmw in the PLAIN FR-Admin corpus). It thus seems to be relatively more specific to the discourse of administrative French.

- (10) *L'autorité administrative statue sur la demande <dans un délai de six mois> à compter du dépôt par l'étranger du dossier complet de cette demande.* (FR-Law)  
 The authority reviews the application **within six months** after the application has been made. (Our translation)

Whereas the pattern is found in sequences wherein the reader is directly addressed in the PL administrative texts, the legislative excerpt (Example 10) is much more impersonal, as the subject of the verb is an abstract entity (*the administrative authority*).

In the specialised texts (FR-LEX, CA-Judgements), we find a number of sequences in which the subject of the verb corresponds very often to an abstract legal or administrative concept. In the administrative corpus, the typical subject of certain patterns corresponds

more often to the law user (thus representing a significant re-orientation of the discourse). This difference suggests what has been called a process of ‘personalisation’ (Turnbull 2018), by which expert legal knowledge is reformulated for a non-expert readership.

For example, one of the most common POS-grams we find in the corpus is <Vous devez/pouvez + V>, involving the second person *vous* and a modal verb expressing either obligation (*devez* = must) or possibility (*pouvez* = can/may). The lexical verb that is introduced in these contexts often expresses an administrative procedure (expressed as a Material or Behavioural process<sup>3</sup>).

(11) <Vous devez écrire> directement au procureur de la République.

**You must** write to the public prosecutor directly (Our translation). (FR-Admin-PL)

(12) <Vous pouvez collecter> vous-même les preuves de ce harcèlement.

**You can** collect evidence of your harassment yourself (Our translation). (FR-Admin-PL)

Similarly to *dans un délai de*, this pattern is usually associated with various types of legal action, albeit not as fixed in terms of the syntactic frame. The modals of both the obligation and possibility vary according to the way in which legal dissemination texts define the user’s legal rights and obligations. As with previous examples of reorientation, the use of the second-person pronoun is part of the communication strategy called ‘conversationalization of public discourse’ (Turnbull 2018), which is performed through the use of direct questions (as seen above) as well as first- or second-person pronouns to reformulate highly abstract legal knowledge and to create a form of dialogue between the institution and the non-expert readership.

### 3.2.2. Summary of the Canadian Plain Language Corpus

The most salient POS-grams and lexical bundles in the summaries of judgements are presented in Table 4.

**Table 4.** Main LxGr patterns in French Canadian summaries.

Pattern	Frequency (pmw)	Freq in General French Corpus	Keyness Score
<La Cour d’Appel> <sup>1</sup>	1161.2	0.4	826.2
<La Cour Suprême a + V + que> <sup>2</sup>	450.8	1.5	183.4
<le droit/pouvoir de + V	450.8	0.3	345.4
<Les juges majoritaires ont + V + que> <sup>3</sup>	778.7	0.16	674.4
<la Charte canadienne des droits et [libertés]> <sup>4</sup>	220.5	0.02	2061.3

<sup>1</sup> Translation: <The Court of Appeal>. <sup>2</sup> Translation: <The Supreme Court + V + that>. <sup>3</sup> Translation: <The majority + V + that>. <sup>4</sup> Translation: <The Canadian Charter of Rights and Freedoms>

The LxGr patterns in the CA-PL-Summaries are mostly Institutional bundles, referring to either institutions (the Court of Appeal), legal actors (judges) or to foundational legal texts: in particular, the Canadian Charter of Rights and Freedoms, which sets out and protects a number of rights and freedoms, as can be seen below.

(13) <La Cour d’appel> a dit partager l’opinion du premier juge. (CA-PL-Summaries)

**The Court of Appeal** agreed with the trial judge. (Official English version)

(14) M. Chhina a fait valoir que son traitement était illégal au regard de <la Charte canadienne des droits et libertés>, qui fait partie de la Constitution du Canada.

Mr. Chhina said that his treatment was illegal under the **Canadian Charter of Rights and Freedoms**, part of Canada’s constitution. (Official English version of the summary by the Supreme Court of Canada)

It can be noted that several of the shorter LxGr patterns presented in Table 4 and found in the CA-PL-Summaries are the same as those found in the CA-Judgements corpus: namely, *la Cour d’appel* (‘the Court of Appeal’) or an extended LxGr pattern of the type *les juges majoritaires ont + V + que* (‘the majority judges + V + that’). These mostly correspond

to key terms of judicial discourse in the Common Law system, which explains why they are used abundantly by the Supreme Court Justices in their judgements. As for the summaries, they have both an encapsulating and explanatory function. As such, they need to explain what the various courts involved in a case have decided and, in particular, what the majority opinion of the judges was; hence, there are frequent references to Courts of Appeal and to the majority judges.

What is interesting is that many of the sequences found through the analysis of POS-grams reveal extended LxGr patterns, often corresponding to projection (expression of engagement through indirect speech).

- (15) <Les juges majoritaires ont affirmé que> la police a porté atteinte aux droits que la Charte garantit à M. Reeves en prenant l'ordinateur sans son consentement et sans mandat. (CA-PL-Summaries)

**The majority said** that the police breached Mr. Reeves' Charter rights by taking the computer without his consent and without a warrant. (English version of the summary by the Supreme Court of Canada)

- (16) <La Cour suprême a confirmé que>, dans une cause criminelle, le doute «raisonnable» doit être fondé sur la preuve et non sur des conjectures.

In a criminal case, 'reasonable' doubt should be based on evidence, not speculation, **the Supreme Court has confirmed**. (English version of the summary by the Supreme Court of Canada)

These examples appear to belong to an extended LxGr pattern for which we suggest the following formula: <Institutional subject + Communicative process (state, confirm) + que (that) + Reporting clause>. This pattern is linked to a key inherent function of the summary genre, which is to report the Supreme Court's decisions by explaining the opinions expressed not only by the Justices but also by the inferior courts. Some authors discuss this in terms of 'discursive heterogeneity' (Preite 2016), referring to the fact that legal popularisation discourse (like all popularisation discourse) is based on a 'primary' specialised discourse that is explicitly mentioned as a legitimising source. This is especially true in the summaries of decisions, as the figure of the judge is at the core of the explanation and elaboration strategies.

### 3.3. Focus on Phraseology: Patterns from Legal Texts Also Found in the PLAIN Corpora

We have seen a certain number of patterns that appear to be specific to certain genres or are found in a certain type of register (legal or popularised text). We now turn to some specific patterns that seem to be of particular interest because they are particularly representative of legalese, and they can also be found in PLAIN legal texts from our corpus.

#### 3.3.1. *Il y a* (There Is/There Are)

The first pattern we want to focus on is the existential or presentational structure *Il y a*, which is usually translated as *There is/are*. Although its keyness score is not extremely high, the pattern caught the authors' attention when analysing the data, as it is used with a specific syntax and is associated with an extended LxGr pattern that has a specific discourse function in the concordances in both legal and plain texts. It furthermore exhibits a higher relative frequency across our corpora as compared to the reference corpus (respectively: FR-Law 125.9 pmw; FR-Judgements 109.3 pmw; FR-Admin-PL: 132.9 pmw vs. FrenchTenTen 34.11 pmw). Some examples of this pattern and its use are set out in Figures 1 and 2.

### LxGr spécifique aux corpus CA JUG, FR LOI: 'Defining (the conditions of) a Legal Term.'

< [Theme] **Il y a** + N (dénomination d'un délit / crime) **lorsque** + [Définition] >

A1) Est qualifié **crime ou délit** flagrant le crime ou le délit qui se commet actuellement, ou qui vient de se commettre. **Il y a** aussi **crime ou délit flagrant lorsque**, dans un temps très voisin de l'action, la personne soupçonnée est poursuivie par la clameur publique, ou est trouvée en possession d'objets ...

A2) Les sanctions qui ne sont pas incompatibles peuvent être cumulées ; des dommages et intérêts peuvent toujours s'y ajouter. **Il y a** **force majeure** en matière contractuelle **lorsqu'**un événement échappant au contrôle du débiteur, qui ne pouvait être raisonnablement prévu lors de la conclusion du contrat...

A3) **Annulation** de l'acte instrumentaire<sup>162</sup> **Il y a lieu à annulation lorsque** l'acte est irrégulièrement dressé, bien que ses énonciations soient exactes.

A4) ... tout mur servant de séparation entre bâtiments jusqu'à l'héberge, ou entre cours et jardins, et même entre enclos dans les champs, est présumé **mitoyen** s'il n'y a titre ou **marque** du contraire. **Il y a** **marque de non-mitoyenneté lorsque** la sommité du mur est droite et à plomb de son parement d'un côté, et présente de l'autre un plan incliné.

A5) La **violence** est une cause de nullité qu'elle ait été exercée par une partie ou par un tiers. **Il y a** également **violence lorsque** une partie, abusant de l'état de dépendance dans lequel se trouve son cocontractant à son égard, obtient de lui un engagement qu'il n'aurait pas souscrit en l'absence d'une telle contrainte ...

Figure 1. Concordances of presentational structure <Il y a + NG (term)> in French specialized legal texts.

< [Theme] **Il y a** + N (dénomination d'un délit / crime) **si / quand** + [Définition] >

B1) Cas d'**abus de confiance**. **Il y a** **abus de confiance quand** une personne s'approprie un bien que lui a confié sa victime. </s><s> Ce bien peut être une somme d'argent

B2) Différence avec vol et escroquerie. L'**abus de confiance** se distingue de **l'escroquerie**. **Il y a** **escroquerie si** l'auteur fait croire qu'il possède un droit sur le bien (par exemple, si l'auteur des faits retire de l'argent sur le compte de la victime avec une fausse procuration)

B3) **Téléservice** **Harcèlement** scolaire - Violences scolaires - Provocation au suicide. **Il y a** **harcèlement scolaire quand** un élève fait subir à un autre, de manière répétée, des propos ou des comportements agressifs.

B4) La simple **tentative** de vol ou de racket suffit pour rendre une plainte recevable. </s><s> **Il y a** **tentative si** l'auteur des faits a commencé à commettre son infraction mais qu'elle a échoué à cause d'un élément indépendant de sa volonté.

B5) Les **violences conjugales** peuvent correspondre à des violences : psychologiques, physiques, sexuelles, ou économiques. </s><s>. **Il y a** **violence conjugale quand** la victime et l'auteur sont dans une relation sentimentale.

Figure 2. Concordances of presentational structure <Il y a + NG (term)> in French PLAIN texts.

In the excerpts from FR-Law and CA-Judgements shown in Figure 1, this pattern is used to define a legal term or the conditions for a legal term to be established, especially a crime or felony. Similarly, in the excerpts from FR-Law and CA-Judgements shown in Figure 2, a similar pattern is used to specify and define a legal term. In the figures, the existential structure "il y a" is shown in black, while the term is presented red and the conjunction introducing a conditional clause or definition is in purple font. Significantly, this is always a term that has just been mentioned in the immediate co-text. Grammatically speaking, the indefinite article is always elided in this construction (before the term), which

is unusual in French: hence, our choice to our focus on this structure. This particular feature allows us to draw up a formula for the LxGr pattern as a whole: <[term in preceding discourse] ... *il y a* + [Ø] + [name of the term] // conditional clause + definition>. Finally, a notable difference in the corpora is that in the PL texts, the clause that introduces the definition or conditions for the crime or felony is *quand* (when) or *si* (if), while the legal texts contain its more formal alternative *lorsque*.

The structure *Il y a* + [Ø *article*] + *definition* happens to be used in other popularisation/dissemination discourses: for example, medical popularisation. The example below is an excerpt from the French National Health Service website explaining the signs of cardiac arrest, and it introduces the term using the same <*il y a* + [Ø]> construction.

(17) <**Il y a arrêt cardiaque si**> : *la victime perd connaissance, tombe et ne réagit pas quand on lui parle ou qu'on la stimule*

**It is a case of cardiac arrest** if the victim is unconscious, falls or shows no reaction when talked to or stimulated (our translation).

Other types of dissemination discourse that use this pattern include promotional discourse or game rules. It appears to be specific to elaboration strategies and definitions of terms in French.

### 3.3.2. Prepositional Cascades

The second pattern we now turn to is often found in specialised legal texts and is especially highly frequent in the FR-Law corpus, although the CA-Judgements corpus also contains some examples as well as variations on the same type of structure. The pattern in question can be seen as an expanded version of the complex prepositional chains mentioned in Section 3.1, as it is composed of a chain of three prepositional phrases. What is surprising is that this highly complex structure can also be found routinely in the texts from the PLAIN corpus. Table 5 displays the frequency (pmw) of this POS-gram in the specialised corpora and their simplified versions as well as an example from each corpus.

**Table 5.** A salient POS-gram across corpora: Prepositional cascades <Prep + N + prep + N + prep>.

Corpus	Relative Frequency (pmw)	Example
FR-LAW	7150	Sont exonérés < <b>de droits de mutation par</b> > décès le conjoint survivant et le partenaire lié au défunt par un pacte civil de solidarité.
FR-Admin-PL	5702	Faire un recours préalable auprès de la MSA < <b>par courrier de préférence en</b> > recommandé avec avis de réception
CA-Judgements	2015	Les juges ordonnent fréquemment à des individus, < <b>à titre de condition à leur mise en liberté sous caution</b> >, d'éviter tout contact avec l'alcool et les drogues.
CA-Summaries	1502	Dans un tel cas, l'accusé peut interjeter appel de la déclaration < <b>de culpabilité pour meurtre au</b> > deuxième degré

What stands out in this table is that, first, the prepositional cascades <Prep + N + prep + N + prep> are particularly prominent not only in the FR-LAW corpus, with a frequency of 7150 pmw, but also in the FR-Admin-PL, with 5702 pmw. Although this is half as frequent in the CA-PL-summaries than in the CA-Judgements corpus, it is also statistically salient in the judicial dissemination corpus. This result suggests that the syntactic complexity that characterises legal language (Bhatia 1983), even after simplification to address a lay audience, still percolates into the plain texts under study. We discuss our results in the following section.

## 4. Discussion

The results of our phraseological exploration of legal French are consistent with other phraseological studies of legal language in other languages and contexts, as they underline the highly intertextual and syntactically complex nature of legal French in both

subcorpora. Our results also suggest differences in the two genres represented here, as the use of complex prepositions and prepositional cascades appears to be more frequent in the legislative FR-Law than in judicial discourse, represented by CA-Judgements, a result which is also consistent with previous research on legal phraseology in other languages, such as Spanish (Pontrandolfo 2021). Of course, the fact that the two legal subcorpora come from different countries, France and Canada, might also account for this difference in the frequency of these prepositional structures, as legal phrasemes are ‘bound to a particular legal system’ and the results should therefore be interpreted with caution (Pontrandolfo 2023). Future research on comparable judicial and legislative corpora in both European-French and French-Canadian would be necessary to examine whether the difference is genre-based or culture-based.

Concerning the PLAIN corpus, our preliminary results show that there are traces of legal phraseology in both dissemination corpora (that is to say, in the Canadian French corpus and the European French corpus). This becomes particularly clear in the case of projecting/reporting structures and complex prepositional chains. This leads us to suggest that the simplification of legal phraseology in French, though evident in both PLAIN corpora, is only achieved to a limited extent: an observation that is consistent with previous findings (Rossetti et al. 2020). There are, however, some LxGr patterns that appear to be characteristic of plain legal French, especially for certain genres. We note the very high frequency of lexical bundles like *De quoi s’agit-il?* (‘What is it?’) or *Où s’adresser* in the FR-Admin-PL corpus, as illustrated again by the example below in which both patterns create a dialogical structure in a text that is explaining what to do in case of harassment.

- (18) <*De quoi s’agit-il ?*> *Le harcèlement est le fait de tenir des propos ou d’avoir des comportements répétés ayant pour but ou effet une dégradation des conditions de vie de la victime. (. . .)*  
<*Où s’adresser ?*> *La victime peut porter plainte contre le ou les auteurs du harcèlement.*  
(FR-Admin-PL)

**What is it?** Harassment is defined as repeated words of behaviours that aim at or result in the deterioration of the victim’s living conditions. (. . .) **Where to get help**  
The victim can complain about the person(s) that are harassing them.

In our preceding analysis, we have seen lexical patterns that either serve to ‘structure’ the text (Halliday’s textual metafunction), to introduce specialised terms (the *‘Il y a’* construction) or draw the law users’ attention to specific points in the text. Thus, we can characterise the overall strategy in the FR-DILA corpus as an attempt to provide readers with heuristic guidance—helping them to navigate around administrative procedures rather than setting out a ‘boiled down’ or simplified version of these procedures.

In the PL summaries from the Supreme Court of Canada, the communicative strategy is quite different: these texts are structured by references to the central figure—the judge—and the judicial institutions, which include the Supreme Court as well as the lower courts of appeal. Structurally, the core phraseology of this genre is oriented towards reporting clauses or to expansions (explanations), both of which involve extensive use of subordinate and embedded clauses. This can be seen in the example below.

- (19) Selon **les juges majoritaires**, cela surcharge le système judiciaire, **qui** consacre plus d’argent à essayer de faire en sorte que des gens pauvres paient leur suramende qu’il en obtiendrait de ceux-ci. <**Les juges majoritaires ont**> fait observer <**qu**>‘une peine est plus efficace si elle est adaptée à la personne.  
**The majority** said this also burdened the justice system, **which** spent more trying to get poor people to pay the surcharge than it would ever get back. **The majority noted that** a sentence works best if it is made for the individual. (Official English version)

One of the notable differences between the two PLAIN corpora is their types of syntactic complexity. Our preliminary analysis suggests that the Fr-PL-Admin corpus is characterised by more nominal complexity, i.e., by the use of prepositional cascades and complex noun groups, while the CA-summary texts appear to contain more clausal complexity, i.e., more subordinate clauses, especially reported speech. In this sense, the

French administrative texts are closer to the legislative texts they are based on and are more complex in terms of information packaging in noun and prepositional phrases. The French Canadian corpus also exhibits more clausal complexity and is thus more ‘elaborate’ in this particular sense of clausal complexity, as can be seen in Example 19, in which we have emphasised relative pronouns as well as binding and linking conjunctions. Generally speaking, it might be expected that ‘simplified discourse’ will, in fact, turn out to be more elaborate (i.e., involve more structural expansions) than the highly codified but also much denser and compact ‘expert discourse’. Indeed, greater clausal and verbal complexity linked to elaboration strategies and unpacking of information has also been shown in other PL discourse: in particular, medical discourse (Gledhill et al. 2019). The results we have set out above may also be due to the influence of the Plain Language Movement in French-speaking Canada, whereas the concept has not been institutionalised as such in France. If the FR-Admin corpus is closer to the FR-Law corpus, it may also be due to the fact that some of the texts published are actually direct quotations, without any simplification, of the original legal text, as suggested by results put forward by Bouyé (2022). Despite these differences in communicative strategies and complexity, our findings regarding ‘plain French’ appear to be consistent with recent research on legal popularisation; as is the case in other online lay-oriented discourse, drafters ‘balance impersonal explanatory strategies with more interpersonal and communicative strategies’ (Diani et al. 2023).

In the first section of this paper, we raised several questions about the nature and distribution of lexico-grammatical (LxGr) patterns. Our first question related to the relative ‘size’ of n-grams (What is the smallest sequence possible or useful to identify meaningful LxGr patterns?). In the data analysis above, we have demonstrated that it is often possible to analyse short n-grams as longer stretches of meaningful text that often correspond to specific discourse functions. The primary example here is the very short sequence *Il y a*, which on its own can only be seen as a short verbal group in French. Out of context, this sequence is so ubiquitous in French that it tells us very little, but once we begin to look at the corpus data, *Il y a* turns out to be a very distinctive definitional routine in both of the main FR corpora we analysed here.

Regarding our more general questions, it is useful to deal with each of them in turn:

- (1) Is it possible to assign a discourse function to random n-grams, assuming that these sequences have been found to be salient in one of the subcorpora? Here, we have demonstrated this as a principle, although we have clearly not shown this positively across a wide range of data.
- (2) Is it possible to identify characteristic LxGr patterns in legal texts (i.e., in non-plain legal texts)? This has been demonstrated using various examples, such as complex nominal groups/prepositional phrases (Section 3.1).
- (3) Is it possible to identify characteristic LxGr patterns of PL in administrative discourse? This has also been demonstrated in relation to turn-taking sequences and procedural constructions, among other examples (Section 3.2).
- (4) Is it possible to establish a difference between generic phraseology (belonging to several ‘genres’) and specific phraseology (patterns that are ‘unique’ or at least more salient in one genre as opposed to all the others)? We have shown that certain constructions (such as the projection structures of reported speech) are ‘generic’ and occur as significant LxGr patterns in all of the corpora we have analysed here. This is especially evident in the two specialised corpora, in which LxGr patterns related to cross-referencing pervade both judicial and legislative corpora. Despite obvious macro-textual differences, the FR-Law and CA-Judgements corpora can be identified as belonging to the legal register as a whole; this is, notably, based on the ‘generic’ patterns related to cross-referencing, one of the features that gives legal language its characteristic ‘legal flavour’ (Maley 1994). Regarding specific LxGr patterns, we have identified a significant sample of these, among the most recognisable ones being the multiple prepositional phrase pattern (associated with FR Law and CA-Judgements)

as well as the *Il y a* definitional pattern (specific to the LEX corpus by use of the conjunction *lorsque* or to the PLAIN corpus by use of the subordinators *quand/si*).

## 5. Conclusions

In this paper, we have tried to characterise the phraseology of two legal genres and of two types of legal popularisation texts in French from France and from Canada. Our results suggest that salient n-grams and LxGr patterns can be identified in both legal French and PL discourse. Some patterns appear to be generic patterns, i.e., they are linked to the legal register as a whole, while others are specific to a genre or text type. In particular, our findings contribute to the characterisation of administrative and legal dissemination in French as phraseological patterns linked to knowledge recontextualisation and elaboration. Traces of highly complex phraseological patterns from legal language can be found in both PLAIN corpora, although they are more salient in European French texts, suggesting that these are more complex (in the sense of 'lexico-grammatically elaborate') than Canadian French texts. There are probably deep-rooted cultural reasons for this (higher expectations placed on French-speaking users of the law/public services, the relatively recent emergence of the plain language movement in France, differences in legal culture as well as practice ...).

Further research is needed to confirm the preliminary observations set out in this article. Possible research perspectives include designing a survey to measure the comprehensibility of plain legal French by obtaining readability measures (self-paced reading or eye-tracking) from ad hoc tasks targeting specific features in legal texts: for example, complex prepositional phrases or passives.

**Author Contributions:** Conceptualisation, M.B. and C.G.; methodology and analysis C.G.; validation, M.B. and C.G.; formal analysis, M.B. and C.G.; investigation, M.B. and C.G.; resources, M.B.; data curation, M.B. and C.G.; writing—original draft preparation, M.B. and C.G.; writing—review and editing, C.G.; supervision, C.G.; project administration, M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Publicly available datasets were analysed in this study. These data can be found here: <https://www.scc-csc.ca/case-dossier/cb/index-fra.aspx> accessed on 2 November 2023; <https://www.service-public.fr/> accessed on 2 November 2023; <https://www.legifrance.gouv.fr/> accessed on 2 November 2023.

**Acknowledgments:** The authors would like to thank the anonymous reviewers and coordinators of this issue.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

det	Determiner
LxGr	Lexico-Grammatical Pattern
N	Noun
NP	Noun Phrase
Num	Number
PL	Plain Language
pmw	Per Million Words
Prep	Preposition



## Notes

- <sup>1</sup> i.e., ‘constructions’ in Goldberg’s 1985 sense of the term, such as the passive, the resultative, etc.
- <sup>2</sup> Complex prepositions involve the embedding of one prepositional phrase within another (such as *for the purpose of this section*). Sometimes, these structures undergo lexicalisation in which the first prepositional group functions as a single preposition (Biel et al. 2015).
- <sup>3</sup> The terms in capital letters relating to Transitivity in the Systemic Functional Linguistics model are borrowed from (Halliday and Matthiessen 2013).

## References

- Asprey, Michele M. 2004. Plain Language around the world. In *Plain Language for Lawyers*, 3rd ed. Sydney: Federation Press.
- Bhatia, Vijay K. 1983. Simplification v. Easification—The Case of Legal Texts. *Applied Linguistics* 4: 42–54. [CrossRef]
- Bhatia, Vijay K. 1994. *Cognitive Structuring in Legislative Provisions*. London: Longman, vol. 1.
- Biel, Lucja. 2015. Phraseological profiles of legislative genres: Complex prepositions as a special case of legal phrasemes in eu law and national law. *Fachsprache* 37: 139–160. [CrossRef]
- Biel, Lucja. 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*. London and New York: Routledge, vol. 10, p. 26.
- Bouyé, Manon. 2022. 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. Ph.D. Thesis, Université Paris-Cité, Paris, France.
- Bouyé, Manon, and Christopher Gledhill. 2019. Disseminating legal language for the general public: A corpus-based study of the discursive strategies used in english and french. In *Langues et Langues Juridiques. Traduction et Traductologie, Didactique et Pédagogie. Colloque International de Bordeaux*. Paris: Institut Francophone pour la Justice et la Démocratie, pp. 349–69.
- Breeze, Ruth. 2017. Giving voice to the law: Speech act verbs in legal academic writing. In *Phraseology in Legal and Institutional Settings*. London: Routledge, pp. 221–39.
- Cavalieri, Silvia. 2018. Broadcasting legal discourse. the popularization of family law through youtube. In *Popularization and Knowledge Mediation in the Law/Popularisierung und Wissensvermittlung im Recht*. Münster: LIT Verlag, pp. 251–70.
- Cacchiani, Silvia. 2018. The voice of the law on gov.uk and justice.gouv.fr: Good value to citizens and institutions? In *Popularization and Knowledge Mediation in the Law*. Edited by Jan Engberg, Silvia Cacchiani, Karin Luttermann and Chiara Preite. Münster: LIT Verlag, pp. 117–48.
- Collette, K., M. P. Benoît Barnet, D. Laporte, F. Pouëch, and B. Rui-Souchon. 2002. *Guide Pratique de la Rédaction Administrative*. Paris: Ministère de la Fonction Publique.
- Coulthard, Malcolm, Alison Johnson, and David Wright. 2016. *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.
- Crystal, David, and D. Davy. 1969. *Investigating English Style: English Language Series*. London and Harlow: Longmans.
- Cutts, Martin. 2008. *Plain English Lexicon*. London: Plain Language Commission.
- Dąbrowski, Andrzej. 2017. Reel justice in the context of teaching legal english as a foreign language. *Glottodidactica* 43: 107–20. [CrossRef]
- des Droits, Défenseur. 2019. *Dématérialisation et Inégalités D'accès aux Services Publics*. Rapport. Paris: Défenseur des Droits.
- Diani, Giuliana. 2018. Popularization of legal knowledge in English and Italian information books for children. In *Popularization and Knowledge Mediation in the Law*. Münster: Lit Verlag, pp. 291–316.
- Diani, Giuliana. 2023. Disseminating legal information on online law forums in english and italian. *Iberica* 46: 299–320. [CrossRef]
- Engberg, Jan, Karin Luttermann, and Silvia Cacchiani. 2018. *Popularization and Knowledge Mediation in the Law/Popularisierung und Wissensvermittlung im Recht* Münster: LIT Verlag, vol. 9.
- Galonnier, Bernard. 1997. Le discours juridique en France et en Angleterre. Convergences et spécificités. *ASP la Revue du GERAS* 15–18: 427–38. [CrossRef]
- Gledhill, Christopher, Hanna Martikainen, Alexandra Mestivier, and Maria Zimina-Poirot. 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* 91–114.
- Gledhill, Christopher, Stéphane Patin, and Maria Zimina. 2017. Lexico-grammaire et textométrie: Identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français. *Corpus* 17. [CrossRef]
- Goźdz-Roszkowski, Stanisław. 2012. Discovering patterns and meanings: Corpus perspectives on phraseology in legal discourse. *Roczniki Humanistyczne* 60: 47–70.
- Goźdz-Roszkowski, Stanisław, and Gianluca Pontrandolfo. 2013. Evaluative patterns in judicial discourse: A corpus-based phraseological perspective on american and italian criminal judgments. *International Journal of Law, Language and Discourse* 3: 9–69.
- Goźdz-Roszkowski, Stanisław, and Gianluca Pontrandolfo. 2017. *Phraseology in Legal and Institutional Settings: A Corpus-Based Interdisciplinary Perspective*. London: Routledge.
- Halliday, Michael Alexander Kirkwood. 1994. Spoken and written modes of meaning. In *Media Texts: Authors and Readers*. Clevedon: Multilingual Matters Ltd., pp. 51–73.

- Halliday, Michael Alexander Kirkwood, and Christian MIM Matthiessen. 2013. *Halliday's Introduction to Functional Grammar*. London: Routledge.
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý, and Vít Suchomel. 2013. The tenten corpus family. Paper presented at 7th international Corpus Linguistics Conference, Lancaster, UK, July 23–26. pp. 125–27.
- Kilgarriff, Adam. 2009. Simple maths for keywords. Paper presented at the Corpus Linguistics Conference 2009 (CL2009), Liverpool, UK, July 20–23. vol. 6.
- Luttermann, Karin, and Jan Engberg. 2018. Vermittlung rechtlichen wissens an kinder und jugendliche im internet und in broschüren. In *Popularization and Knowledge Mediation in the Law/Popularisierung und Wissensvermittlung in Recht*. Münster: LIT Verlag, pp. 85–115.
- Maley, Yon. 1994. *The Language of the Law*. London: Longman, vol. 1, pp. 11–50.
- Masson, Michael E. J., and Mary Anne Waldron. 1994. Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology* 8: 67–85. [CrossRef]
- Plain English Campaign. 2022. How to Write in Plain English. In *Plain English Campaign Drafting Resources*. New Mills: Plain English Campaign. Available online: [www.plainenglish.co.uk/free-guides.html](http://www.plainenglish.co.uk/free-guides.html) (accessed on 21 May 2022).
- Pontrandolfo, Gianluca. 2021. National and eu judicial phraseology under the magnifying glass: A corpus-assisted analysis of complex prepositions in spanish. *Perspectives* 29: 260–77. [CrossRef]
- Pontrandolfo, Gianluca. 2023. The importance of being patterned. In *Handbook of Terminology*. Amsterdam: John Benjamins Publishing Company, pp. 124–150.
- Preite, Chiara. 2016. La vulgarisation des termes juridiques et la construction d'un savoir («que» faire) chez le grand public. *REPÈRES-DORIF* 10: 1–9.
- Preite, Chiara. 2018. Stratégies dialogiques et transmission du savoir juridique dans le site du ministère de la justice française. In *Popularization and Knowledge Mediation in the Law/Popularisierung und Wissensvermittlung im Recht*. Münster: LIT Verlag, vol. 9, p. 149.
- Rossetti, Alessandra, Patrick Cadwell, and Sharon O'Brien. 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*. Berlin and Heidelberg: Springer, pp. 699–711.
- Tiersma, Peter M. 2000. *Legal Language*. Chicago: University of Chicago Press.
- Turnbull, Judith. 2018. Communicating and recontextualizing legal advice online in English. In *Popularization and Knowledge Mediation in the Law/Popularisierung und Wissensvermittlung im Recht*. Münster: LIT Verlag, pp. 201–22.
- Williams, Christopher. 2004. Legal English and plain language: An introduction. *ESP Across Cultures* 1: 111–24.
- Williams, Christopher. 2015. Changing with the Times: The Evolution of Plain Language in the Legal Sphere. *Revista Alicantina de Estudios Ingleses* 28: 183. [CrossRef]
- Williams, Christopher. 2022. *The Impact of Plain Language on Legal English in the United Kingdom*. Abingdon: Taylor & Francis.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.