



Disseminating legal language for the general public: a corpus-based study of the discursive strategies used in English and French

Manon Bouyé, Christopher Gledhill

► To cite this version:

Manon Bouyé, Christopher Gledhill. Disseminating legal language for the general public: a corpus-based study of the discursive strategies used in English and French. *Langues et Langages juridiques. Traduction et traductologie, didactique et pédagogie. Colloque international de Bordeaux, Jun 2019, Bordeaux, France. pp.349-369. hal-04050459*

HAL Id: hal-04050459

<https://u-paris.hal.science/hal-04050459>

Submitted on 29 Jun 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.



Disseminating legal language for the general public: a corpus-based study of the discursive strategies used in English and French

Manon Bouyé & Christopher Gledhill¹

Résumé : Dans cet article, nous examinons la façon dont la langue des textes législatifs est reformulée par les organisations et institutions juridico-administratives afin que les justiciables puissent comprendre leurs droits et prendre des décisions éclairées. Nous mettons en avant une approche descriptive de l'analyse de textes, à l'aide des outils les plus récents de la linguistique de corpus. Nous proposons une méthode systématique pour comparer un corpus de textes législatifs, réputés complexes et rédigés par des experts, à un corpus de textes dits simples, destinés aux non-experts, afin de mettre au jour les stratégies de reformulation et de simplification aux niveaux macro- et micro-textuels. Cette étude comporte une dimension comparative français-anglais qui interroge les différences et similarités entre contextes culturels.

Mots-clés: médiation juridique, linguistique de corpus comparée, phraséologie, discours de simplification, plain language.

Abstract: In this paper, we study how the language of legislative texts is reformulated by organizations and government agencies, so that non-expert citizens can understand and use these texts efficiently to know their rights and make informed decisions. We emphasize a descriptive approach to text analysis, using the latest methods of corpus linguistics. We set out in particular a systematic method for comparing a corpus of texts that are reputedly complex (the expert-written corpus here called LEGAL) and a corpus of texts that are reputedly simple (or 'simplified', here known as PLAIN). These texts have specific properties that set them apart from other registers and genres, a feature of language variation that is often overlooked when talking about 'plain language'. We thus pay particular attention to the divergent or similar rhetorical strategies of reformulation and simplification of legal language at macro- and micro-textual levels across languages (French and in English).

Key-words: expert to non-expert legal communication, comparative corpus linguistics, phraseology of simplified discourse, plain language

¹ Université Paris-Diderot.

I. Introduction

Communication between specialized legal institutions and the general public has long been an object of study and there have been many proposals to implement ‘plain language’ from legal experts, political authorities and researchers. The aim of this paper is to present some of the discourse strategies used in the dissemination of legal texts for non-experts from the point of view of descriptive linguistics. To do this, we compare a corpus of British and French legal texts, considered to be highly technical and complex, with a corpus of texts that are claimed to express legal concepts and processes clearly for a non-expert readership (i.e. British and French citizens). We also explore the lexicogrammatical characteristics of texts that are claimed to have undergone simplification following the drafting recommendations published by authorities or organizations which advocate plain legal communication.

II. Legal language and its dissemination

A. Clarifying legal language for non-experts: plain language, dissemination and simplification.

The complexity of legal language, be it the language of private contracts, wills, laws, or judicial decisions, has often been pointed out by consumer organizations and legal experts alike in both English- and French-speaking countries². This has led to the publication of recommendations for expert legal drafters by institutions and organizations around the world, such as the Plain English Campaign in the UK or the European Union. The overall goal of such recommendations is to make legal documents more readable and easier to use for non-experts, so that they can have full knowledge of their rights and thus make more informed, safer legal decisions.

In many countries, proponents of clearer legal and administrative texts have relied on the concept of plain language, which, like other controlled languages, is based on a set of lexical and grammatical rules³. In English-speaking countries in particular, criticism of the complexity of legal language has led to a Plain Language Movement. In each of those countries, non-governmental organizations (NGOs) have worked with government agencies and departments to help them redraft complex legal documents. Some organizations, like the Citizens Advice Bureaux (CAB) in the United

2 Michele ASPREY, David Neil TURNER, *Plain language for lawyers*, Sydney : Federation Press, 1996, p. 14.

3 Christopher GLEDHILL, Hanna MARTIKAINEN et al., « Towards a linguistic definition of ‘Simplified Medical English’: applying textometric analysis to Cochrane medical abstracts and their plain language versions. », in M -Ch JULLION, L.M. CLOUET & I. CENNAMO (dir.), *Les institutions et les médias aujourd’hui : de l’analyse du discours à la traduction*, Lingue Culture Mediazioni, Editioni universitarie di Lettere Economia Diritto, 2019, vol. 11, pp 91-114.

Kingdom, take it upon themselves to disseminate public information about the law, and claim to use plain language in the content they publish. The rules prescribed in the style guides notably include avoiding the passive voice and addressing the law users directly; as for lexical items, the use of archaic terms and jargon is discouraged.

In France, the dissemination of legal specialized knowledge is not primarily taken up by NGOs but by government agencies such as the *Direction de l'Information Légale et Administrative*, a centralized institutional body since 2010. Some researchers have also produced recommendations to be used by institutions which publish material for the general public⁴; they include very similar recommendations to that of English, mentioned above. In both France and the UK, these recommendations have been designed to clarify the characteristics of legal language, especially statutes, and which we now briefly present.

B. Legal language: a brief overview

Although both judicial and courtroom discourses have been abundantly described in the literature, for the purpose of this study, we focus on legislative discourse, and in particular on the specific genre of the statute (i.e. a text that embodies law). Legislative discourse is arguably the most common type of legal discourse to which non-expert users are exposed daily; it is also the type of discourse whose simplified or ‘clear’ versions are the most directly accessible online to build a corpus of texts.

What are the characteristics of legal statutes that make them complex for non-experts? Apart from the high degree of prior specialized knowledge which is necessary to understand legal terminology, legislative texts are characterized by very specific syntactic, lexical and discursive features. As is well known, the French and British legal systems are fundamentally different, the first being typical of continental legal systems and the latter being the prototypical common law system. However, due to the historical influence between legal discourse of French and English, we suggest here that the features set out for one language can be equally relevant for both. These similarities are especially relevant when we consider that both languages can share a ‘discourse’, that is to say the implementation of language in a particular context by a particular community (as in ‘the written discourse of medical research’, ‘the oral discourse of sports commentary’, etc.).

Generally speaking, legal discourse in both English and French can be characterized by abstraction (as opposed to concreteness) and depersonalization (downplaying of personal engagement). In order for the law to be applicable to a range of contexts, legal statutes use a form of abstract language,

⁴ Karine COLLETTE, Marie-Paule BENOÎT BARNET *et al.*, *Guide pratique de la rédaction administrative*, Ministère de la fonction publique et de la réforme de l’État, en ligne sur <http://www.dusa.gouv.fr/cosla/index.htm>, 2002.

which may come close to the language of logic in the case of English. In addition, we assume that abstract language is more difficult to understand than concrete language (according to studies on language acquisition⁵). Sample 1 gives an example of this:

(1) A person A discriminates against a woman if, in the period of 26 weeks beginning with the day on which she gives birth, A treats her unfavourably because she has given birth. (Equality Act 2010)

Sample 1 is abstract because it involves : a) impersonal nouns ('person A'), b) multiple complex clause embedding and post-modification ('beginning with the day on which...') and c) circumspect legal terms rather than everyday lexical items ('treated unfavourably' = 'treated unfairly').

(2) La discrimination est punie de trois ans d'emprisonnement et de 45 000 euros d'amende lorsqu'elle consiste :

1) A refuser la fourniture d'un bien ou d'un service

2) A entraver l'exercice normal d'une activité économique quelconque (Code pénal, Article 225-2)

Sample 2 shares many of the features that we can see in Sample 1. It also displays another characteristic of legal discourse which is common to both languages: a tendency to use grammatical metaphor⁶, or nominalizations, i.e. the use of nouns (here *fourniture*, *exercice*), to refer to processes that can be expressed by verbs (*refuser*, *exercer*). Space precludes a discussion of this in any detail here, but it is worth noting that for many analysts, this phenomenon is considered to be a reliable indicator of complexity, as more lexical information can be packed in a noun phrase than in a clause⁷.

In brief, the relative complexity of legal language can be associated with a specific configuration of linguistic features: syntactic expansion (embedding, post-modification), lexical density (including nominalization), and highly specific rhetorical strategies.

C. Research goals

This study has three main research goals:

1) Our first aim is to identify the discourse strategies of simplification as they emerge in reputedly simplified texts. More specifically, we attempt to determine the changes in stance and textual voice which take place during the passage from a legal text, mostly read by experts, to an informative text aimed at non-expert readers. We address these issues by examining two comparable and representative corpora in both English and French.

⁵ Thomas FRANÇOIS, Cédric FAIRON, « Les apports du TAL à la lisibilité du français langue étrangère. » *Traitement Automatique des Langues*, vol. 54, no. 1, 2013.

⁶ Michael HALLIDAY, Christian MATTHIENSEN, *An Introduction to Functional Grammar*, London: Edward Arnold, 2004, p. 710.

⁷ Michael HALLIDAY, Christian MATTHIENSEN, *idem*.

2) A second aim is to examine the extent to which the recommendations made by style guides are actually followed in practice: for the purposes of this paper, we focus on English. If there is a significant gap between the recommendations and actual practice, we formulate a hypothesis as to why authors do not abide by the prescribed rules (Section IV.B).

3) Finally, this preliminary study has a cross-linguistic dimension, as our corpus is composed of texts in English (UK) and French (France). Can we bring to light similar linguistic reformulation strategies in the dissemination of legal knowledge? Conversely, how do these discourse strategies differ across two different languages and cultural contexts?

In the following section, we set out how we address these issues by using the descriptive methods of corpus linguistics as well as the conceptual framework of Systemic Functional Grammar.

III. Corpus and methods

A. A bilingual corpus

In order to explore reformulation strategies and their impact on lexis and grammar, our first step is to build a corpus of texts from French and British parliamentary websites, respectively *Légifrance* and *legislation.gov.uk*. This ‘expert’ corpus (which we term LEGAL) is divided into two subcorpora, each containing 20 whole statute texts for UK English and about 3 000 law articles for French, for a total of 750,000 words in each corpus. For the ‘non-expert’ (= PLAIN) corpus, we chose texts published by organizations which in both countries have historically been seen as the main mediators of legal information. These texts were collected in two archives of texts published between 2017 and 2019 and described as follows:

- The EN-CAB subcorpus, made up of 773 texts published online by the Citizens Advice Bureau (CAB), a British NGO which has a long history of disseminating legal knowledge for the general public. The total number of tokens in this subcorpus is 711 935;
- The FR-DILA subcorpus, made up of 712 texts drafted by the *Direction Générale de l'Information Légale et Administrative* (a department of the French Prime Minister’s office), which, like Citizens Advice in the UK, is one of the most popular legal information websites for law users in France. It contains 751 020 word tokens.

In the first instance, we limit our analysis to a preliminary case study on a sample of texts defining the concept of discrimination in both British and French law (as explained in Section III). This analysis is then extended to the whole corpus, as discussed in Section IV, to see if the recurrent phraseology that can be observed in the PLAIN subcorpora corresponds to identifiable

popularization strategies.

B. Methodological framework and tools

As mentioned above, the analytical framework we use here is Systemic Functional Grammar (SFG), a model which provides a structured approach to describing lexicogrammatical features and rhetorical functions across different text types. For the purposes of this study, we concentrate on just one key feature of this model: process type, i.e. the semantic relationship expressed by a verb and its arguments within the clause. In our analysis, we examine the following questions: what is the proportion of semantic process types in plain vs. legal texts? Does legal language return a larger proportion of one type of semantic process, as compared to those used for the dissemination of legal knowledge? Does the intended readership affect the proportion of the types of semantic processes in the text?

First, in order to compare semantic processes in both corpora quantitatively, we extracted basic verb forms using the corpus analysis program SketchEngine⁸, for the French and English texts. The extraction was performed using the Corpus Query Language based on the part-of-speech tags from SketchEngine. The meaning of the verbs in the present tense in the different subcorpora was then hand-coded in a spreadsheet. We then used the typology from SFG to compare the semantic processes expressed by verbs in the LEGAL and PLAIN samples. The main categories of semantic processes include: Material processes, which have to do with doing and happening (*discriminate against, treat someone*); Mental processes, which have to do with perception (*see, hear*), cognition (*know, understand*), affection (*like, love*), and desire (*hope, want*); and Relational processes, which are used to link two or more entities (often used in definitions, for instance, *une discrimination constitue une distinction*). Relational processes are often expressed by stative verbs such as *seem, become, or have, own, possess*⁹. Once the verbs had been hand-coded, the normalized frequencies of the different types of processes were then calculated. For this case study, we also used verb extraction to identify the passive forms more easily. This allows us to compare the language actually used by legal advice organizations with their published recommendations. We present and discuss the results of the case study in Sections III.

After the case study, we examined the whole PLAIN corpus (CAB and DILA) by extracting ngrams, i.e. recurring sequences of n words or ‘lexical bundles’¹⁰. The analysis of ngrams allows us to establish the core phraseology of a text or a corpus of texts. In this paper, phraseology is seen as

8 Adam KILGARRIFF BAISA, Jan Vít BUŠTA *et al*, « The Sketch Engine: ten years on », *Lexicography*, 2014, vol. 1, no 1, p. 7-36.

9 David BANKS, « La Linguistique Systémique Fonctionnelle: une approche sémantique et sociale », *hal-archives-ouvertes.fr*, 2011.

10 Douglas BIBER, Susan CONRAD, « Lexical bundles in conversation and academic prose », *Language and Computers*, 1999, vol. 26, p. 181-190.

the identification of recurring sequences of words, which generally also operate as whole units semantically – they also express a regular, predictable discourse function or share a general meaning, such as “Organize the text” or “Define a term”. For each chain of *n* words, we selected the two most frequent ngrams from the corpus, and then used the concordance tools to study how each sequence is used in context in the DILA and CAB subcorpora (Section V).

IV. Case study: explaining and clarifying the legal concept of discrimination to non-experts

We begin our study of the discourse strategies of law popularization by looking at a smaller subset of our corpus: those legislative texts which define the concept of discrimination. For English, this sample includes the Equality Act of 2010, the most recent law defining discrimination. We then selected ten texts from the CAB subcorpus, published between 2018 and 2019, which were explicitly based on the Equality Act. For French, we collected excerpts from the *Code Pénal* and other Codes which dealt with discrimination, as well as texts from the DILA subcorpus which explain what discrimination is and what legal actions can be taken against it. We chose discrimination because it is relatively equivalent in British and French law, and because in both languages, the PLAIN (dissemination) texts were explicitly based on the legislative texts, thus allowing us to directly compare the differences between a complex text and its simplified version in both contexts.

A. Semantic processes in statutes and their simplification

We first turn to the results of semantic process analysis, i.e. the types of process expressed by verbs in our two sample corpora. What can they reveal about the passage from a highly complex legal discourse to informative discourse for non-experts? The results for the distribution of each type of semantic process in each subcorpus are displayed in Table 1.

Table 1. Distribution of semantic processes (%) in each subcorpus

	Normalized frequency of Relational processes (%)	Average normalized frequency of Mental processes (%)	Normalized frequency of Material processes (%)	Number of verbs
CAB sample	43.7	18.3	38	1233
DILA sample	43.4	13.9	41.9	762
Equality Act sample	75.7	5.9	18.4	239

Légifrance sample	50.8	17.8	31.4	242
-------------------	------	------	------	-----

Table 1 shows that Relational processes are distinctly overrepresented in legislative texts (the FR-LEGAL subcorpus), with 50.8% of Relational processes, and even more overrepresented in the EN-LEGAL subcorpus, with a little over 75%, whereas in the dissemination corpora (CAB, DILA) there is a greater variety of types of semantic process. This difference suggests that legislative texts, both in English and French, especially in laws on discrimination, define and specify the concept of discrimination and, using identifying or attributive Relational processes, set out precisely the characteristics and the persons who are protected by the law.

On the other hand, the CAB and DILA subcorpora seem characterized by much greater variety in the types of semantic processes, especially a higher proportion of Material processes in both French (41%) and English (38%). This shows that the dissemination of legal knowledge consists of a ‘recontextualization’ of specialized knowledge¹¹. The legal concept of discrimination as defined by the law is recontextualized by using references to situations that law-users can encounter daily in their social life, at work or in the public sphere, through the use of Mental processes (e.g. *The driver tells you to stop breast-feeding or get off the bus*) or Material processes (e.g. *Le syndicat agit en votre nom*).

The discrepancy in the distribution of process types is notably greater in English, especially for Mental processes, as there are three times as many Mental processes in the CAB sample than in the EN-LAW sample, while the proportion of Mental processes in the DILA and FR-LAW samples only differ by about 4%. We discuss the possible reason for this difference later.

B. Plain language drafting and going against the rules: the case of the passive

As previously mentioned, one of the recommendations prescribed by Plain Language guidelines is to avoid passive structures, which are typically considered more complex. In actual dissemination texts, however, we observe that this recommendation is not always respected. This is especially true for English and texts from the CAB corpus, as illustrated here:

Example (3)

Expert version	Plain version
<i>A person A <u>discriminates</u> against another B if, because of a protected characteristic, A <u>treats</u> B less favourably than A <u>treats</u> or <u>would treat</u> others.</i>	<i>Direct discrimination is when <u>you're treated</u> differently and worse than someone else for certain reasons. (...) As well as sex discrimination, you <u>could be discriminated</u></i>

11 Maurizio GOTTI, « Reformulation and recontextualization in popularization discourse », *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos*, 2014, no. 27, p.15-34.

In Example 3, the expert version in the left column contains no passives, unlike the ‘plain’ version which systematically uses it. Why did the CAB drafters use the passive here? In this example, it alters the information structure from the legislative text (cf. the active *A treats B*) in order to place the second-person pronoun, referring to the law user, not later on in the clause (Rheme) but towards the beginning (in Thematic position)¹²; this is called thematicization.

This example shows that the legal dissemination and rewriting of legislative texts (which we suggest corresponds to a recontextualization of legal concepts) goes hand in hand with a reorientation of syntax, here from active to passive. The law user is placed at the centre of the dissemination discourse, in accordance with plain language principles, but paradoxically this also entails a re-orientation of the verb group (introduction of the Relational process ‘to be’ + participle). The study of semantic processes and information structure has shown us that the dissemination of legal knowledge and reformulation of complex texts are achieved through a recontextualization process which not only consists of an explanation of legal terminology, but of a thorough syntactic and semantic reorientation. We now go on to examine how this reflected not just in one sample, but in the whole PLAIN corpus.

V. Corpus study: is there a phraseology of simplification?

We now extend our study to a broader exploration of our two PLAIN corpora. This section presents our preliminary findings on the phraseology of legal knowledge dissemination in French and in English.

A. Phraseology, and the discourse functions of lexical bundles

To study phraseology, we use a methodology which has now become standard among corpus linguists¹³, who identify ‘chunks’, ‘extended collocations’ or ‘lexico-grammatical patterns’ as the first step in establishing a systematic analysis of specific types of discourse. First, we extracted recurring sequences of 4 to 7 words from the CAB and DILA corpora, using AntConc¹⁴. To guarantee an optimal distribution of different lexical bundles (not just sequences repeated in one text), we selected the two most frequent sequence of n words that occurred in at least 15% of each

¹² Michael HALLIDAY, Christian MATTHIENSEN, op. cit. (n. 6), pp. 89-103.

¹³ Susan CONRAD, Douglas BIBER, « The frequency and use of lexical bundles in conversation and academic prose », *Lexicographica*, 2005.

¹⁴ Laurence ANTHONY, « Concordancing with AntConc: An introduction to tools and techniques in corpus linguistics », *JACET Newsletter*, 2006, vol. 155, p. 20-85.

dissemination subcorpora (i.e. in at least 100 texts). We then associated these significant ngrams with specific discourse or rhetorical functions, as displayed in Tables 2 and 3.

As can be seen, the discourse functions of different lexical bundles are often very specific to a particular context, although in theory there should be fewer discourse functions than actual realizations of these functions (i.e. for each function, there may be several variable forms). In addition, it is interesting to note that the discourse functions displayed in Table 2 and 3 can themselves be classified according to a broader typology: 1) **referential** expressions, i.e. expressions that refer to elements of daily life or of the legal process; 2) **interpersonal** expressions, which express engagement towards the reader on the part of the drafter, or the drafter's point of view or attitude, including modality; and 3) **metatextual** lexical bundles, which encompass meta-comments on the text itself and intertextual expressions, for example segments which describe the organization of the text or which direct the reader to other texts¹⁵.

Table 2 – Most salient 4-to-7-word ngrams in the EN-CAB subcorpus.

Ngram	Normalized frequency (per million of words)	Discourse function	Example
You may be able to	605	Express a possibility of action for the reader	<i><u>You may be able to</u> take legal action for a breach of one or more human rights under the Human Rights Act 1998.</i>
This is called X	386	Associate a previous notion with a named technical term	<i>The right to claim benefits depends on the terms on which you've been allowed to enter the UK. <u>This is called</u> your immigration status.</i>
This page tells you more about	273	Organize the text, announce the topic	<i><u>This page tells you more</u> about the appeals process you should follow.</i>
You don't have to	163	Express optional actions, the absence of obligation	<i><u>You don't have to</u> wait for the outcome of an appeal before you re-apply for benefits.</i>
It's a good	150	Suggest an action	<i><u>It's a good idea</u> to ask your</i>

15 These functions are ultimately derived from Halliday's three metafunctions (Interpersonal, Textual, Ideational). Michael HALLIDAY, Christian MATTHIESSEN, *op. cit.* (n.6), p.85.

idea to			<i>council if they can help you find a mediator.</i>
If you need more help	140	Direct the reader to an adviser	<i>Contact the Citizens Advice consumer <u>if you need more help</u> - a trained adviser can give you advice over the phone.</i>
Equality and Human Rights Commission	133	Provide the technical name for an institution	<i>You may be able to refer your complaint to the <u>Equality and Human Rights Commission</u>.</i>
Read this page to find out more	117	[Hypertext reference to another page]	<i><u>Read this page to find out more</u> about harassment.</i>

Table 3 - Most salient 4-to-7-word ngrams in the FR-DILA subcorpus.

Ngram	Normalized frequency (per million of words)	Discourse function	Example
dans un délai de	506	Express a time reference	<i>La plainte doit être déposée <u>dans un délai de 6 ans</u> après la découverte des faits.</i>
par recommandé avec avis de réception	353	Refer to an element of the legal process	<i>Cette décision vous est adressée <u>par lettre recommandée avec avis de réception</u>.</i>
de quoi s'agit-il ?	350	Define a term or concept (in the context of an imaginary dialogue)	<i><u>De quoi s'agit-il ?</u> L'adoption crée un lien de filiation entre l'adoptant et l'adopté.</i>
à partir de la + event	286	Express a time reference (starting point)	<i>Ce recours doit être formé devant la cour d'appel dans le délai d'1 mois <u>à partir de la</u> signification du jugement.</i>
c'est-à-dire	266	Define or introduce a term	<i>Toute femme enceinte peut</i>

		explicitly	<i>décider d'accoucher anonymement, <u>c'est-à-dire</u> sous X.</i>
tribunal de grande instance	259	Refer to an element of the legal process	<i>Toutefois, le <u>tribunal de grande instance</u> peut accorder une dérogation.</i>
il s'agit de/d'	225	1) Introduce a term 2) Refer to a term	<i>1) S'il y a eu pénétration, <u>il s'agit d'un viol</u>. 2) S'il <u>s'agit d'une</u> contravention de la 1^{ère} à la 4^e classe, le juge renvoie mineur devant le tribunal de police.</i>
pour vous aider à remplir le formulaire	185	Direct the reader to another text	<i><u>Pour vous aider à remplir le formulaire</u> : Notice relative à la demande de tutelle sur mineur</i>

In the next section, we compare the functions of the ngrams in the English and French subcorpora.

B. Results: interpersonality, metatextuality and reference in the CAB and DILA corpora

According to the results above, a striking feature of the discourse of dissemination of legislative texts is its metatextuality, in both languages. In the four most salient ngrams in the EN-CAB corpus, two have a metatextual function and point to the self-referential nature of the dissemination texts. The two metatextual ngrams (underlined in the examples) however do not have the same discourse function. The most frequent of the two, 'This is called' is used to define a term, as shown in Example 6, and is usually followed by a noun phrase (in bold in the example). The second metatextual ngram 'This page tells you' is followed not by a noun phrase but by a clause (in bold in the example), and is a discourse organizer, which here announces the topic of the page in question, as shown in Example 7.

(6) *Most people who come to the UK have permission to enter the country and spend time here.
This is called **lawful presence**.*

(7) *This page tells you more about **when the police can disclose information about someone who has been convicted of child abuse offences.***

Although both ngrams start with THIS and both are classified as metatextual segments, they do not follow the same lexicogrammatical pattern, nor are they used at the same moment in the text. The ngram <This page tells you> is found at the beginning of a text, while <This is called>, as shown in Example 6, appears anywhere in the text – after one or more sentences which describe a legal concept or process – and concludes the description of the concept by giving the specialized term in question. Those two ngrams thus have very specific metatextual functions. The last ngram <Read this page to find out> has a similar metafunction as <This page tells you>, and can be considered a variant.

The FR-DILA subcorpus is also characterized by several metatextual segments. The ngram <Pour vous aider à remplir le formulaire> in the FR-DILA subcorpus has both an intertextual function, as it links to another text (usually a hyperlink to a pdf text), and an instructional function, as it incites the reader to take the next step in a legal process by filling out the appropriate form. It is highly formulaic and recurrent in the DILA corpus.

Three other ngrams from the list are metatextual: <De quoi s'agit-il ?>, <il s'agit> and <c'est-à-dire>. They have the same functions as the two metatextual segments in the CAB corpus. <De quoi s'agit-il?> is also a discourse organizer and is found at the beginning of the texts (like <This page tells you more about>), as it introduces the legal concept or process that the text talks about, as seen in Table 3. As for <c'est-à-dire>, it serves as a reformulation device to elaborate or define terms, as shown in Example 8.

(8) *La décision de refus peut être explicite, c'est-à-dire écrite et motivée, ou implicite, si le maire n'a pas répondu dans le délai d'1 mois.*

Some lexical bundles can have more than one function. <Il s'agit> for instance, can be used either to introduce and explain a new term, as seen in the first example from Table 3, or to refer to a previously mentioned term, as seen in the second example.

The abundant use of metatextual and reformulation segments, which characterize the two subcorpora, is consistent with previous research on popularization¹⁶, and with the observations of Askehave & Nielsen¹⁷, who note the tendency for online texts to favour two modes of reading (traditional 'linear reading' and 'non-linear navigation' which makes use of new cohesive devices and hypertext links). However, our results also suggest two distinct tones adopted by the authors in French and in English, as we will now try to show.

C. Two discursive voices for two cultural contexts

16 Chiara PREITE, « La vulgarisation des termes juridiques et la construction d'un savoir («que» faire) chez le grand public », *REPÈRES-DORIF*, 2016, vol. 10, pp. 1-9.

17 Inger ASKEHAVE, Anne Ellerup Nielsen. « Digital genres: a challenge to traditional genre theory », *Information Technology & People*, 2005, vol. 18, no 2, 120–141

A closer analysis of these ngrams in context suggests two distinct discursive strategies.

In the English subcorpus, the textual voice, which corresponds to the rhetorical image that is built by the writer in discourse (i.e. in a context which requires a cooperative interlocutor who shares the same communicative goals), allows for greater reader identification¹⁸. It appears to be characterized by a form of orality and informality.

Thus several particularly salient ngrams from the CAB corpus displayed in Table 2 can be said to have a strong interpersonal dimension. Only the segment <Equality and Human Rights Commission> has a referential function and refers to a technical term, while the most statistically salient ngrams <You may be able to> (605 per billion of words) and its variant <You might be able to> (190 per billion of words), has an interpersonal function. This reveals a certain stance on the part of the authors, who use hedging to signal to the reader that the clause content – usually linked to legal action or its results – is not entirely certain.

(8) *You may be able to challenge the decision.*

(9) *You might be able to get your pension sooner if you're retiring due to ill health.*

Here, according to the contexts, modality is used by the authors of the dissemination texts to express a probability for the reader to be able to perform an action or obtain a result during a legal process. It is notable that in the extended context of the n-gram, we find a variable set of items which nevertheless share a predictable meaning, in this case either a) verbs expressing legal or paralegal procedures, or b) conditional elliptical phrases such as ‘do this if’ which set out the conditions in which the legal procedure may proceed. We term this productive (but also predictable) extended phraseology a ‘lexicogrammatical pattern’. Note also in these examples the use of modality and of MAY and MIGHT as a form of rhetorical precaution on the part of the authors, who cannot predict with certainty whether the content of the text in question will apply to the law user. This is due to one of the paradoxical characteristics of the law, which is intended to be universal but also depends on individual cases (especially in the common law tradition)¹⁹, so the drafters use MAY to imply that it might also not be the case.

The highly distinctive expression <It's a good idea to> also has an interpersonal dimension, as seen in Example 10.

(10) *If you make the complaint verbally, it's a good idea to make notes beforehand of what you want to say as it's easy to get distracted if you're upset. (CAB corpus)*

The appreciative adjective ‘good’ combined with the contraction of the verb form BE (*it's...*), expresses an informal suggestion on the part of the drafters as to what actions the reader should take at such steps of a legal process. This shows that legal dissemination genres not only popularize legal

18 Dominique MAINGUENEAU, « L’ethos discursif et le défi du Web » *Itinéraires. Littérature, textes, cultures*, 2016, vol. 3, p.96.

19 John Peter GIBBONS, *Language and the Law*, Routledge, 2014.

knowledge, but also have an instructional function, like recipes or instruction manuals: they guide the law user through the legal process²⁰. Here, they do it in a relaxed, informal mode which is close to spoken discourse.

The study of ngrams thus shows an increased interpersonal dimension in the English (CAB) corpus. By contrast, the French DILA corpus seems to have a much less interpersonal dimension. Out of the 8 ngrams listed in Table 3, only one, <Pour vous aider à remplir le formulaire> addresses the reader. Four of the most statistically salient lexical bundles in Table 3 have a referential function, as they serve to introduce a time reference, for example the deadline to appeal a judicial decision, or to refer to an element of the legal process, like the concerned court (cf. <tribunal de grande instance>) or the preferred way in which the law user has to contact or will be contacted by the institutions (<par recommandé avec avis de réception>). The three other ngrams on the list have a metatextual function. Although, <Pour vous aider à remplir le formulaire> directly addresses the reader, the use of the infinitive gives it a more procedural dimension.

These results are not exhaustive, but they do represent the most salient segments in the PLAIN corpus. On the basis of this initial observation, we suggest that the typical tone in the English dissemination corpus is more relaxed, and characterized by hedging, while the tone in the French dissemination corpus is more instructional and is focused on guiding the reader through the legal process, in a more formal tone.

VI. Discussion

A. Discourse strategies of legal dissemination in French and English

What do these preliminary results show about the differences between legal language and the discourse of dissemination or ‘plain language’? First, we find that the dissemination of legal knowledge is achieved through a reformulation process which affects both information structure and grammatical structure simultaneously (thus the choice between Material / Mental process or Attributive / Identifying process at the level of the verb group also has an impact on the choice between active / passive orientation above the level of the clause). The simplification of legislative texts consists of a recontextualization of specialized legal knowledge, where recontextualization takes place at a macro-textual level (with discourse organizers) and is in turn reflected in grammar and information packaging at a micro-textual level. When the prescribed recommendations for plain

20 Chiara PREITE, *op. cit.* (n. 16).

language are not respected by authors, it can usually be traced back to recontextualization strategies that help the reader apprehend complex legal processes or concepts.

Our results also show a clear difference in discourse strategies used to disseminate legal knowledge and to give instructions to non-experts in French and English. The texts from the DILA dissemination corpus tend to reveal a preference for a formal and procedural tone in its phraseology, whereas the texts from the CAB corpus tend to be closer to oral, informal language, with a more varied use of semantic processes to recontextualize legal concepts. One conclusion that we could draw from this is that both UK and French texts correspond in many ways to the respective socio-cultural expectations: in the UK positive politeness and informality are seen as being part of the discourse of consumer-oriented plain language, whereas in France, it is perhaps considered more trustworthy to employ technically precise and stylistically neutral language.

As far as lexico-grammatical characteristics are concerned, it can be noted that the difference in the distribution of semantic processes between the DILA and FR-LAW corpora suggests that French popularizers stay closer to their sources (legislative texts) than the authors in the CAB corpus. We assume that this is because the *Direction Générale de l'Information Légale et Administrative*, which depends on the French Prime Minister's office, is much closer to the legislative institution organizationally. This hypothesis is supported by the fact that every text from the DILA subcorpus explicitly refers to the articles in the legislative texts on which it is based, through the use of hyperlinks to the legislative text in question at the end. In other words, the institutional discourse appears to stay closer to the legislative texts it is based on. This is not the case for the CAB texts (although some of them, like the texts explaining discrimination, explicitly refer to the legal genre). There is of course an important cultural element at play. In the UK, Citizens Advice is an organization which is present at a local level; its tone is more informal, and more cautious towards its content, hence the central place of modality in its phraseology, a feature called *hedging* that can also be found in other discourses, such as scientific writing.²¹ We suggest therefore that the CAB subcorpus is more than a reformulation of the legislative texts, but a complete redrafting in plain language. The reason for a strong interpersonal dimension in the CAB corpus can also be found in the recommendations for plain language. The difference in strategies can thus be said to stem from a different culture of legal dissemination in France and in the UK, suggesting that the importance of the “territoriality”²² of legal genres and their popularization has an influence on the lexicogrammatical characteristics of plain texts.

21 Ken HYLAND, « Writing without conviction? Hedging in science research articles », *Applied linguistics*, 1996, vol. 17, no 4, p. 433-454.

22 Sylvie MONJEAN-DECAUDIN, « Territorialité et extraterritorialité de la traduction du droit », *Meta: Journal des traducteurs/Meta: Translators' Journal*, 2010, vol. 55, no 4, p. 693-711.

B. Limits and perspectives

Finally, we would like to discuss some interesting problems that arise when considering our corpus data. First, it is clearly important to recognize that the CAB subcorpus is the product of a non-governmental organization while the DILA subcorpus represents the same institutional framework as the original legal texts. There is therefore an obvious difference of status in the discursive sources chosen to build our corpus, and it is hardly surprising therefore that we observe a discontinuity of phraseology in the English corpora, and at the same time a continuum of discourse practice between French law and legal dissemination. In future studies, we intend to examine text types that have no direct institutional link to legal authorities, for example the equivalent of CAB in other French-speaking countries.

An important factor that we have not had space to develop in this paper is the difference between the legal systems in France and the UK, the first being based on civil law, while the second is based on common law. While such contextual differences cannot be dismissed, we would suggest that differences in phraseology can only be partly imputed in part to differences in legal culture: we would rather emphasize the role of ‘discourse strategies’ here, i.e. differences in the culture of mediation which are probably just as important to realize legal dissemination and which likely have a considerable impact on the discourse strategies discussed here.

VII. Conclusion

In this paper, we have tried to characterize the discourse of legal knowledge dissemination in French and English. The limits of this study include a methodological bias introduced by the discursive sources for our corpus, which are not of the same institutional nature.

These issues notwithstanding, our results do however suggest that there are similar strategies of recontextualization and syntactic reorientation (moving from abstruse legal text to clear, informative dissemination texts). These processes can be observed in the use of metatextual expressions, a finding which is consistent with previous research. One novel feature which we would claim to have observed here is that although legal dissemination texts have an instructional function in both French and in English, the study of phraseological patterns and semantic processes also reveals a different textual voice and different ways to address the non-expert reader according to the cultural context, with a preference in the French corpus for a formal and procedural tone, as opposed to an oral, interpersonal dimension in English. As mentioned above, we suggest that this is linked to deep cultural differences (and expectations) in legal communication.

Future research will focus on the comparison of legal communication from equivalent institutional sources. After studying legislative discourse, we propose to examine judicial discourse and its dissemination, notably by studying Supreme Court judgments and their plain language summaries.