

Uma análise preliminar da microdiacronia nos códigos brasileiros: O Código de Águas e o Código do Consumidor

A Preliminary Analysis of Micro-diachrony in Brazilian Codes: the Código de Águas and the Código do Consumidor

Lucia Almeida Ferrari

Universidade Federal de Minas Gerais (UFMG) | Belo Horizonte | MG | BR
ferrari.lu@gmail.com
<https://orcid.org/0000-0002-9855-0646>

Nayane Araujo

Universidade Federal de Minas Gerais (UFMG) | Belo Horizonte | MG | BR
nanayanearaujo@gmail.com
<https://orcid.org/0009-0008-3225-284X>

Luisa Ramos de Oliveira

Universidade Federal de Minas Gerais (UFMG) | Belo Horizonte | MG | BR
luisaramos@gmail.com
<https://orcid.org/0009-0002-0650-4907>

Carolina Godoi de Faria Marques

Universidade Federal de Minas Gerais (UFMG) | Belo Horizonte | MG | BR
carolgodoi350@gmail.com
<https://orcid.org/0009-0004-3927-4428>

Resumo: Este estudo analisa casos de variação linguística em textos legislativos brasileiros, com base nos dados do *corpus* LEX-BR-lus (Ferrari; Marques, 2022). Foram examinados o Código de Águas (Brasil, 1934) e o Código de Defesa do Consumidor (Brasil, 1990), focando na evolução das formas verbais e nominais por seu papel central na estrutura sintática dos textos. A pesquisa se insere no campo da microdiacronia da linguagem jurídica, que investiga mudanças em períodos curtos, considerando o formalismo característico da legislação e as tendências de simplificação. O *corpus* utilizado contém 755 leis federais, totalizando mais de três milhões de palavras. Os textos analisados foram escolhidos por critérios como época de promulgação e número de modificações. O Código de Águas sofreu poucas alterações, enquanto o Código de Defesa do Consumidor passou por onze mudanças. A análise envolveu a marcação e a categorização das formas verbais, comparando anotações automáticas e verificações manuais. Os resultados indicam tendências de mudança na redação das normas, sugerindo um possível afastamento do modelo arcaico. Estudos futuros expandirão essa análise para um número maior de textos, incorporando métodos estatísticos e ferramentas computacionais para validar os achados preliminares.

Palavras-chave: microdiacronia; leis federais brasileiras; LEX-BR-lus; verbos.



Abstract: This study analyzes linguistic variation examples in Brazilian legislative texts based on data from the LEX-BR-Ius *corpus* (Ferrari; Marques, 2022). The research examines the *Código de Águas* (Brasil, 1934) and the *Código de Defesa do Consumidor* (Brasil, 1990), focusing on the evolution of verbal and nominal forms due to their central role in the syntactic structure of the texts. The study is situated within the field of micro-diachrony in legal language, which investigates changes over short periods, considering both the formalism characteristic of legislation and trends toward simplification. The *corpus* comprises 755 Brazilian statutory legal texts, totaling over three million words. The texts analyzed were selected based on criteria such as the date of enactment and the number of modifications. The *Código de Águas* has undergone a few changes, whereas the *Código de Defesa do Consumidor* has been amended eleven times. The analysis involved tagging and categorizing verbal forms and comparing automatic annotations with manual verification. The results indicate trends in textual modifications, suggesting a possible shift from archaic models. Future studies will expand this analysis to a broader set of texts, incorporating statistical methods and computational tools to validate these preliminary findings.

Keywords: micro-diachrony; Brazilian statutory federal laws; LEX-BR-Ius; verbs.

1 Introdução

Neste artigo apresentaremos os resultados preliminares de um projeto piloto realizado a partir dos dados do LEX-BR-Ius (Ferrari; Marques, 2022), um *corpus* representativo das leis federais brasileiras.

O LEX-BR-Ius compilou 755 textos legais em vigor extraídos do site Portal da Legislação¹ com um total de 3.300.289 palavras. Cada texto foi devidamente processado em três versões, além dos respectivos cabeçalhos contendo seus metadados: (a) uma versão limpa em formato .txt para possíveis processamentos futuros; (b) uma versão em formato Modest XML, com etiquetas estruturais criadas para o projeto segundo a proposta de Hardie (2014); (c) uma versão com etiquetagem morfossintática com o *parser* PALAVRAS (Bick, 2000, 2014), todas com codificação UTF-8.

¹ BRASIL. Casa Civil. *Subchefia para Assuntos Jurídicos*. Legislação – Presidência da República. Disponível em: <https://www4.planalto.gov.br/legislacao/>. Acesso em: 30 jan. 2025.

Na pesquisa que aqui é apresentada, dois textos específicos, a dizer, o Código de Águas, Decreto nº 24.643, de 10 de julho de 1934 (Brasil, 1934), doravante CA, e o Código de Defesa do Consumidor, Lei nº 8.078, de 11 de setembro de 1990 (Brasil, 1990), doravante CDC, foram submetidos a uma análise quantitativa e qualitativa de tipo microdiacrônica, que averiguou a variação das formas verbais e nominais de todos os verbos ao longo do tempo, observando as diferentes modificações pelas quais os dois textos legislativos passaram. Optou-se por perfazer uma busca partindo justamente dos verbos por estes serem o centro discursivo dos textos, que, nos casos das normas legislativas, apresentam construções extremamente complexas, ricas em subordinadas e listas de direitos e deveres a serem cumpridos, como mostra o Exemplo 1 abaixo:

- (1) Art. 20. O fornecedor de serviços responde pelos vícios de qualidade que os tornem impróprios ao consumo ou lhes diminuam o valor, assim como por aqueles decorrentes da disparidade com as indicações constantes da oferta ou mensagem publicitária, podendo o consumidor exigir, alternativamente e à sua escolha:
- I - a reexecução dos serviços, sem custo adicional e quando cabível;
 - II - a restituição imediata da quantia paga, monetariamente atualizada, sem prejuízo de eventuais perdas e danos;
 - III - o abatimento proporcional do preço (Brasil, 1990, grifo nosso).

Pode-se observar a oração principal, construída em volta do verbo “responder” seguida por uma série de subordinadas, emaranhadas entre si, com os verbos “tornar”, “diminuir”, “poder”, “exigir”, além da lista de nominalizações.

Como o objetivo do trabalho é verificar se, e em qual medida, haveria mudança ao longo do tempo na linguagem utilizada nas leis, partiu-se dos verbos, em suas formas verbais e nominais, de maneira a traçar um quadro mais amplo. Ao mesmo tempo, dada a necessidade de uma leitura atenta de cada parágrafo, foi efetuada uma análise qualitativa de vários fenômenos, como a posição do sujeito e de algumas construções peculiares.

A análise se desdobrou na classificação e na verificação manual de um total de 2.308 ocorrências de formas verbais e nominais dos verbos com seus relativos contextos.

O artigo está assim dividido: após uma introdução geral sobre o conceito de microdiacronia na linguagem legal, a seção de metodologia descreve o *corpus* utilizado e as etapas da análise. Em seguida são apresentados os resultados obtidos e sua discussão, seguidos pelas considerações finais.

2 Microdiacronia na linguagem legal

Linguagem jurídica ou legal é um termo “guarda-chuva” que engloba uma série de áreas distintas, desde a língua utilizada em processos e nas várias esferas do judiciário e aquelas das leis, até as peculiaridades analisadas nas perícias forenses, entre outras. O tema tem despertado, há um certo tempo, o interesse dos linguistas, mas também de outras áreas, dada sua componente multidisciplinar: em sua análise é importante levar em conta as características históricas e culturais, além das tradições jurídicas de cada país (Tiersma, 2000; Coulthard, Johnson, 2007, 2010; Richard, 2018; Goźdz-Roszkowski, 2011, 2012). As análises linguísticas têm se valido

de diversas perspectivas, entre as quais a linguística de *corpus* para quantificar as análises e identificar padrões morfossintáticos e lexicais (Pontrandolfo, 2012; Goźdz-Roszkowski; Pontrandolfo, 2015; Biel, 2018; Giampieri, 2018; Goźdz-Roszkowski, 2021; Ferrari; Cunha, 2022).

A literatura geralmente descreve a linguagem jurídica, e especificamente aquela utilizada nas leis dos diferentes países, como marcada por alto grau de formalismo, utilização de terminologia técnica, muitos arcaísmos, grande complexidade sintática e redundância (Cortelazzo, 1997; Tiersma, 2000; Dahlman, 2006; Rodrigues, 2007; Ferraz Junior, 2014; Williams, 2004, 2020, 2023). Trabalhos recentes têm buscado verificar se realmente as leis, e a linguagem jurídica em geral, têm mantido um grande apego às tradições e ao conservadorismo, apesar de todo o movimento internacional para a simplificação da linguagem legal, mais conhecido como *plain language movement* (Adler, 2012) e da própria evolução da língua e influências de outros idiomas (Adler, 2012; Cortelazzo, 2012; Visconti, 2012, 2017; Canavese, 2019). Dentre tais trabalhos citam-se Canavese (2019), que compilou um *corpus* de leis do italiano suíço para análises em microdiacronia e traça um panorama amplo sobre os trabalhos na área, e Giampieri (2024), que investigou em diacronia os latinismos nas decisões da Suprema Corte do Reino Unido.

Em referência ao presente trabalho, esclarece-se que as leis brasileiras são modificadas com uma frequência variável.² Após sua promulgação, podem ficar em vigor por anos, serem rapidamente substituídas por novas normas, assim como terem trechos mais ou menos amplos modificados inúmeras vezes. Algumas normas são extremamente recentes, outras bem mais antigas, datando da metade do século XIX (como é o caso da Lei nº 556, de 25 de junho de 1850, o Código Comercial, da época do Império), mas ainda em vigor até hoje. Trata-se, portanto, de dados muito interessantes para uma análise em microdiacronia. Como será descrito abaixo, foram selecionadas não somente as leis em vigor no momento da coleta, mas também assinaladas, com um aparato de anotações, as diferentes modificações pelas quais os textos passaram, permitindo assim o detalhamento de eventuais mudanças ao longo do tempo. À primeira vista, a linguagem das leis brasileiras ainda soa como muito complexa, às vezes indecifrável e com construções arcaicas. As perguntas de pesquisa giram justamente em torno desse ponto: a) Até que ponto tal arcaísmo é devido a leis cuja redação é antiga e ainda estão em vigor? b) Haveria mudanças ao longo do tempo ou se foi replicando o mesmo modelo de redação do começo da República até os dias de hoje?

Na próxima seção serão detalhados o *corpus* utilizado e a metodologia adotada para tentar responder a tais questões.

3 Metodologia

Será apresentado brevemente, nesta seção, o *corpus* utilizado para a pesquisa e, em específico, serão detalhados os dois textos escolhidos: o Código de Águas e o Código de Defesa do Consumidor. Em seguida, será descrito de que forma foram separados e analisados os dados da pesquisa.

² A cada dia as leis federais podem passar por modificações, seja sob forma de alterações parciais ou totais de seus artigos ou supressões que visam atualizar sua abrangência. As normas podem, portanto, apresentar diferentes modificações, que são fruto de tais mudanças.

O *corpus* LEX-BR-lus (Ferrari; Marques, 2022) foi compilado entre janeiro de 2022 e janeiro de 2024. Seu objetivo é fornecer dados representativos das leis federais brasileiras em vigor no período da compilação, devidamente limpos e processados para análises linguísticas. Trabalhos anteriores e em preparação detalharam as escolhas metodológicas na limpeza e no processamento dos textos (e.g.: Ferrari; Marques, 2022; Marques, 2023). O *corpus* é dividido em diferentes seções que seguem a divisão legislativa adotada pelo Portal da Legislação, de onde os dados foram extraídos.

Tabela 1 – Seções do *corpus* LEX-BR-lus e suas estatísticas

Seções	Textos em vigor	Textos selecionados	Nº palavras	Nº palavras menor texto	Nº palavras maior texto	Média nº palavras
Códigos	17	13	711.952	12.225	167.416	54.765,50
Constituição	1	1	97.082	N/A	N/A	N/A
Emendas à Constituição	114	48	50.241	43	12.015	1.046,69
Estatutos	18	11	125.356	2.370	35.241	11.396,00
Leis complementares	192	43	182.768	90	35.837	4.250,42
Leis ordinárias	13.491	639	2.132.890	37	86.611	3.337,86
TOTAL	13.833	755	3.300.289	N/A	N/A	N/A

Fonte: Elaboração própria

Pode-se notar que há uma grande disparidade no tamanho de cada lei e no número de leis de cada seção. Entre as decisões metodológicas, aquela de manter a integridade dos textos foi uma das mais importantes, pois a estrutura de cada um inclui diferentes partes e até temas que qualquer recorte teria enviesado (Sinclair, 2004; Biber, 1993). Já a seleção de alguns textos e não de outros seguiu o critério de uso (Barbera; Onesti, 2009; Onesti, 2011; Ferrari; Marques, 2022): cada uma das quase 14.000 leis em vigor na época da coleta foi verificada no JusBrasil (2023), o maior portal agregador de dados jurídicos brasileiros, sendo escolhidas aquelas mais frequentemente utilizadas em processos e citações legais.

Entre as seções do *corpus*, do ponto de vista do uso e da própria importância legislativa, a Constituição, os Códigos e os Estatutos são os mais influentes. Especificamente, os Códigos estabelecem as normas e princípios que norteiam amplas áreas do mundo legislativo, como o Direito Penal e o Tributário. Foram selecionados para este estudo piloto os dois Códigos citados anteriormente devido a três fatores principais: (a) o tamanho relativamente reduzido e similar, mas não excessivamente pequeno dos textos (12.232 palavras³ no CA e 12.495 palavras no CDC), o que permitiu uma análise manual, mais adequada para um piloto, na sua comparação e possibilitou também vislumbrar possíveis etapas e pesquisas futuras; (b) a época de promulgação das duas normas, suficientemente distantes entre si (1934 para o CA e 1990 para o CDC), para possibilitar diferenças linguísticas relevantes; e (c) o número de modificações pelas quais as leis passaram: pequeno (duas no total) no CA e numeroso (onze no total) no CDC.

³ Foi utilizado aqui o termo palavra no sentido de *token*, seguindo a contagem feita pelo Notepad ++, onde os textos foram processados.

O Código de Águas, instituído pelo Decreto nº 24.643, regula a classificação e a utilização das águas no Brasil. Ele foi promulgado em 10 de julho de 1934, durante um período em que o país passava pela transição de uma economia agrária para uma economia industrial. Esse processo de industrialização acelerada demandava um aumento na produção de energia elétrica, o que levou o governo a adotar medidas que facilitassem e impulsionassem a geração de energia hidrelétrica (Cetesb, 2023). O preâmbulo do decreto destaca essa necessidade, visando à criação de “uma legislação adequada que, de acordo com a tendência atual, permita ao poder público controlar e incentivar o aproveitamento industrial das águas” (Brasil, 1934). Além disso, o texto defende que “em particular, a energia hidráulica exige medidas que facilitem e garantam seu aproveitamento racional” (Brasil, 1934).

Nesse contexto, e considerando a inexistência de uma legislação específica sobre o tema até então, foi publicado o Código de Águas. O documento é composto por 205 artigos, totalizando 12.232 palavras, dentre as quais foram identificadas 1.263 instâncias verbais. Apesar da criação de leis complementares posteriormente, o texto da lei passou por apenas duas alterações (em outubro de 1941 e em abril de 1975) e, com exceção dos artigos modificados, manteve-se em sua maior parte fiel à redação original de 1934.

O Código de Defesa do Consumidor, por seu turno, foi promulgado em 11 de setembro de 1990, tendo entrado em vigor cento e oitenta dias após a sua publicação, que ocorreu no dia 12 de setembro do mesmo ano. Ele contém 12.280 palavras, 119 artigos e onze modificações (1993, 1994, 1995, 1996, 1999, 2008, 2009, 2012, 2015, 2017 e 2021). De acordo com Guglinski (2019):

a necessidade de uma cooperação internacional após a 2ª Guerra Mundial, visando especialmente a (sic) reconstrução da Europa no pós-guerra, bem como a natural expansão do mercado, ocasionada pelo regime capitalista, proporcionaram o crescimento e a criação de diversos organismos com vistas à proteção e defesa do consumidor na Europa e em outros países.

Guglinski (2019) cita, por exemplo, a Carta dos Consumidores, redigida em 1976 pela Comissão dos Consumidores, originada a partir da Organização Europeia de Cooperação Econômica (OECE), que veio a ser o primeiro documento oficial na Europa a versar sobre Direito do Consumidor, e destaca o Código Civil alemão (*Bürgerliches Gesetzbuch*, ou BGB), em vigor desde 1º de janeiro de 1900, como sendo uma das inspirações na redação do Código do Consumidor Brasileiro.

Ainda de acordo com o especialista, um novo modelo de direito do consumidor surgiu na década de 1960, nos EUA, com o movimento do consumerismo, pautado no consumo consciente. Nesse contexto, o consumidor passou a ser reconhecido como um sujeito de direitos específicos e foram-lhe atribuídos direitos fundamentais (Guglinski, 2019).

Outro marco importante internacionalmente foi o estabelecimento pela Organização das Nações Unidas (ONU), em 1985, “do princípio da vulnerabilidade do consumidor, reconhecendo-o como a parte mais fraca na relação de consumo, e tornando-o merecedor de tutela jurídica específica, exemplo que foi seguido pela legislação consumerista Brasileira” (Guglinski, 2019).

O consumo no Brasil intensificou-se nos anos de 1930, com a industrialização do país – época em que o Estado já demonstrava forte intervenção na economia. Antes da Constituição de 1988, a defesa do consumidor ganhou destaque com a Lei nº 7.347/85, Lei da Ação Civil Pública (Brasil, 1985), que visava à proteção dos interesses difusos da sociedade e à criação do Conselho Nacional de Defesa do Consumidor, também em 1985.

O CDC, instituído pela Lei nº 8.078/90, é um microsistema baseado em princípios constitucionais. Na Constituição de 1988, o direito do consumidor é reconhecido no art. 5º, XXXII, que prevê que “o Estado promoverá, na forma da lei, a defesa do consumidor” (Brasil, 1988), e a ordem econômica, no art. 170, prioriza justiça social e dignidade. Além disso, ele é principiológico, uma vez que visa a abranger todas as situações envolvendo o consumo, sem, no entanto, especificar cada caso, diferentemente de leis casuísticas.

Buscando acompanhar a evolução das relações de consumo, o CDC passou por muitas modificações desde a sua promulgação. Uma das modificações mais significativas quantitativamente foi a realizada pela Lei nº 14.181, de 2021 (Brasil, 2021), que incluiu dois novos capítulos, ambos tratando do tema de superendividamento – prevenção, tratamento e conciliação.

3.1 Etapas da análise

Como dito acima, os textos foram processados em diferentes versões. Para a finalidade da presente pesquisa, utilizamos a versão anotada em Modest XML (Hardie, 2014) e aquela anotada com o *parser* PALAVRAS (Bick, 2000, 2014).

Na versão em Modest XML (Hardie, 2014), foi criado um conjunto de etiquetas próprias para o projeto. As etiquetas marcam partes relevantes do texto (como “abertura”, “artigo”, “ementa”, “norma”, “promulgação”) assinalando, entre outros fatores, as modificações pelas quais o texto passou, com supressão ou introdução de trechos diversos, e a indicação do ano em que isso ocorreu, conforme ilustrado na Imagem 1.⁴

⁴ No Portal da Legislação, cada norma legal apresenta as partes de leis não mais em vigor como trechos tachados. A etiqueta “tachado”, portanto, contorna os trechos anteriores à modificação. É possível ver, na Imagem 1, que a mesma alínea c é repetida: a versão entre as etiquetas de “tachado” é de 1934, enquanto a versão sem marcação de etiqueta corresponde à versão atual, que é de 1941. A etiqueta “modificacao” permite identificar justamente quando, e com qual ato legislativo, a norma foi modificada. Para o sistema completo de etiquetas utilizado no *corpus* e seus critérios, veja-se Marques (2023).

Imagem 1 – *Printscreen* de um trecho do CA com marcação textual em Modest XML (Hardie, 2014)

```
<artigo>
Art. 144. O Serviço de Águas do Departamento Nacional de
Produção Mineral do Ministério da Agricultura, é o órgão
competente do Governo Federal para: a) proceder ao
estudo e avaliação de energia hidráulica do território
nacional; b) examinar e instruir técnica e
administrativamente os pedidos de concessão ou
autorização para a utilização da energia hidráulica e
para produção, transmissão, transformação e distribuição
da energia hidro-elétrica;
<tachado> c) regulamentar e fiscalizar de modo especial
e permanente o serviço de produção, transmissão,
transformação de energia hydro-electrica: </tachado>
c) fiscalizar a produção, a transmissão, a transformação
e a distribuição de energia hidro-elétrica;
<modificacao> (Redação dada pelo Decreto-lei nº 3.763,
de 25.10.1941) </modificacao>
d) exercer todas as atribuições que lhe forem conferidas
por este Código e seu regulamento.
</artigo>
```

Fonte: Elaboração própria

Já a versão com anotação morfossintática serviu para conferência das análises efetuadas manualmente e para verificar a acurácia do anotador. Os resultados e as tabelas de uma análise multidimensional⁵ implementada no *corpus* inteiro (Marques, 2023; Marques et al., 2025) também completaram o material utilizado para o cotejo das várias ocorrências selecionadas.

O primeiro passo foi efetuar uma varredura manual do CA e do CDC em busca de todas as formas verbais e nominais dos verbos. Cada *token* selecionado foi inserido em uma tabela no Microsoft Excel juntamente com seu contexto de ocorrência. Em colunas separadas indicou-se sua forma no infinitivo e anotaram-se as seguintes informações: pessoa, modo e tempo verbal, voz, as diversas formas de apassivação, eventuais verbos modalizadores e auxiliares, além dos predicativos ligados ao verbo *ser* e da posição dos sujeitos de cada oração.

Como dito acima, as ocorrências foram conferidas seja com a versão anotada em *Part of Speech* do *corpus* com o etiquetador PALAVRAS (Bick, 2000, 2014), seja com os resultados da análise multidimensional. Quando houve dúvidas nas classificações foi consultado um manual de gramática tradicional (Cegalla, 2011) e os casos foram discutidos pelo grupo de trabalho.

Para cada ocorrência foi anotado o ano da modificação, de maneira a efetuar o devido cotejo por período. As Tabelas 2 e 3 apresentam os dados extraídos dos dois textos e que foram analisados nesta pesquisa.

⁵ A análise multidimensional (AMD), proposta por Biber (1988), é uma metodologia baseada em *corpus* amplamente empregada no estudo da variação linguística. Ela possibilita tanto a descrição linguística quanto situacionalmente de diferentes registros, escritos ou orais (Biber 1988, 1995; Cao; Xiao, 2013; Berber Sardinha; Kauffmann; Acunzo, 2014). A AMD permite identificar padrões de coocorrência léxico-gramaticais a partir de cálculos estatísticos que, quando interpretados funcionalmente, revelam dimensões de variação (Biber; Conrad, 2009). Essas dimensões permitem classificar, comparar e descrever os registros analisados em detalhe. Para mais informações, ver Biber (1988) e Berber Sardinha e Pinto (2019). Uma AMD foi implementada em todo o *corpus* em um trabalho recente e seus dados serviram também para comparação do presente cotejo: para informações ver Marques et al. (2025).

Tabela 2 – Número de palavras por período e ocorrências analisadas do Código de Águas

Ano de modificação	N. de palavras	Ocorrências analisadas
1934 (promulgação)	11.963	1.236
1941	234	24
1975	35	2
TOTAL	12.232	1.262

Fonte: Elaboração própria

Tabela 3 – Número de palavras por período e ocorrências analisadas do Código do Direito do Consumidor

Ano de modificação	N. de palavras	Ocorrências analisadas
1990 (promulgação)	9333	873
1993	100	7
1994	56	7
1995	153	15
1996	26	1
1999	42	2
2008	58	6
2009	68	6
2012	31	2
2015	67	8
2017	133	18
2021	2428	201
TOTAL	12495	1146

Fonte: Elaboração própria

Uma vez completadas as tabelas de cada código com todas as informações, utilizando-se as operações da própria planilha Excel, foi efetuado o cotejo das análises por tipos e somadas todas as ocorrências, agrupando assim os dados.

Procedeu-se, em seguida, à sua normalização: foram agrupados e somados os trechos de texto legal dos diferentes períodos históricos. Para viabilizar a comparação entre os dados advindos de trechos de tamanhos tão diversos, como visto nas Tabelas 2 e 3 acima, e dado o número relativamente baixo de instâncias verbais analisadas em alguns períodos, todos os dados foram normalizados em termos de porcentagem. Para facilitar a visualização e compreender melhor possíveis tendências e padrões, para cada resultado foram gerados seus relativos gráficos.

No caso do CA há somente duas modificações após a promulgação do texto (1934), com um baixo número de palavras, além de pouquíssimas ocorrências significativas em tais trechos. Estipulou-se que a comparação com o CDC se faria levando-se em conta como ano-base 1934 somadas às (poucas) ocorrências de 1941 e de 1975. No texto, se fará referência ao CA como representante de 1934. Já no caso do CDC, os maiores trechos modificados e com ocorrências relevantes, além do ano de promulgação (1990), foram encontrados no ano de 2021, com quase $\frac{1}{4}$ do número de palavras do período principal. Os resultados que serão apresentados a seguir, portanto, se farão comparando os períodos de 1934 (CA), 1990 e 2021 (CDC).

Algumas observações não foram de tipo quantitativo, mas serão relatadas abaixo, pois são interessantes e podem servir para buscas futuras no *corpus* inteiro.

Por fim, para a validação estatística dos dados, a fim de verificar se os resultados são estatisticamente significantes, foram realizados os testes indicados por Brezina (2018), a dizer: qui-quadrado (χ^2) (Pearson, 1900),⁶ o V de Cramér (Cramér, 1946)⁷ e cálculo do valor-p (Fisher, 1925).⁸

4 Resultados e discussão

Os dois Códigos analisados possuem, do ponto de vista textual, uma estrutura similar, característica de todas as normas legislativas e descrita nos manuais de redação legislativa (Brasil, 2006, 2018). Há as partes que definem os objetos e contextos que serão deliberados nas disposições normativas, as descrições das diferentes situações de aplicação da norma e as disposições que regulam as medidas necessárias para a implementação da lei.

As análises levaram em conta os seguintes aspectos gerais: (i) *type* verbal; (ii) tipo de predicativo; (iii) vozes verbais; (iv) tipo de passivas; (v) formas verbo-nominais; (vi) modos e tempos verbais; (vii) verbos modais.

Isso esclarecido, na apresentação dos resultados das análises efetuadas, será em primeiro lugar apresentado o léxico utilizado. Pode-se ver, nas Tabelas 4 e 5, os vinte verbos utilizados com maior frequência nos dois textos.

⁶ O teste qui-quadrado avalia a associação entre duas variáveis categóricas. Por convenção é assumido um nível de significância mínimo de 5% ($\alpha = 0,05$). O objetivo é rejeitar a hipótese nula, que assume que não há relação entre as variáveis (Brezina, 2018).

⁷ O V de Cramér mede a força da associação entre as variáveis, variando de 0 (nenhuma associação) a 1 (associação perfeita), calculada com base nos graus de liberdade e no intervalo de confiança. Se o intervalo não incluir zero, isso reforça que há uma associação significativa (Brezina, 2018).

⁸ O valor-p indica a probabilidade de um dado ser estatisticamente (in)significante. Valores inferiores a 0,05 são considerados significantes, enquanto valores iguais ou superiores a esse corte não oferecem evidências suficientes para rejeitar a hipótese nula, sendo considerados sem significância estatística (Brezina, 2018).

Tabela 4 – *Types* dos primeiros 20 verbos mais utilizados no CA e CDC e sua porcentagem de uso em relação ao total de verbos utilizados

CA		CDC	
Type verbal	% de uso	Type verbal	% de uso
ser	6,96	ser	6,89
fazer	3,68	prever	3,23
ter	2,66	tratar	2,27
considerar	1,64	dispor	1,92
tratar	1,48	ter	1,57
obrigar	1,40	aplicar	1,40
pertencer	1,40	legitimar	1,40
estabelecer	1,33	fazer	1,31
exigir	1,33	haver	1,13
haver	1,33	adequar	1,05
conferir	1,25	assegurar	0,96
destinar	1,25	considerar	0,96
conceder	1,17	julgar	0,96
dar	1,17	referir	0,96
dispor	0,93	responder	0,96
permitir	0,93	apresentar	0,87
determinar	0,86	informar	0,87
referir	0,86	obrigar	0,87
caber	0,78	pagar	0,87
servir	0,78	promover	0,87

Fonte: Elaboração própria

Comparando-os com a literatura, alguns resultados são esperados, como a altíssima presença dos verbos *ser* (em todos os casos aqui analisados foi utilizado como verbo de ligação), *ter* e *haver* (Biderman, 2001; Davies; Ferreira, 2006; Rodrigues et al., 2022), enquanto outros são próprios da linguagem setorial das normas legislativas, como os Exemplos 2 e 3 abaixo:

- (2) As informações de que **trata** este artigo, nos produtos refrigerados oferecidos ao consumidor, serão gravadas de forma indelével (Brasil, 1990, cap. V, sec. II, art. 31, par. ún. grifo nosso).
- (3) A competência da União se estende as águas de que **trata** o art. 40, nº II (Brasil, 1934, cap. VII, art. 64, par. ún. grifo nosso).

Destes vinte verbos mais frequentes nos dois Códigos, pode-se ver que alguns se repetem, mesmo que com porcentagens diferentes em relação ao total dos verbos. No CA foram encontrados uma série de verbos que, em construções particulares, se referem especificamente ao assunto abordado na lei, nas amplas seções que definem o âmbito de aplicação da norma, como *fazer* e *pertencer*, conforme os Exemplos 4 e 5 abaixo:

- (4) As correntes de que se fazem os lagos e lagoas navegáveis ou fluviáveis serão determinadas pelo exame de peritos (Brasil, 1934, art. 2, § 2º, grifo nosso).
- (5) As ilhas ou ilhotas, que se formarem no álveo de uma corrente, pertencem ao domínio público, no caso das águas públicas, e ao domínio particular, no caso das águas comuns ou particulares (Brasil, 1934, art. 23, cap. V, grifo nosso).

Já no CDC, o tema em volta do qual giram os verbos mais frequentes é a própria norma legal, como na construção que exemplificamos em (6):

- (6) **Aplicam-se a este artigo**, no que couber, as mesmas regras enunciadas no artigo anterior e as do parágrafo único do art. 22 deste código (Brasil, 1990, art. 44, § 2º, grifo nosso).

São diferenças sutis, e provavelmente devidas ao assunto abordado em cada lei, mas é algo que despertou nossa curiosidade, fazendo-nos questionar se isso seria devido a uma mudança na estrutura interna dos dois textos. De fato, no CA, de 1934, os trechos de definições dos âmbitos de aplicação da lei são muito longos e detalhados, enquanto as disposições normativas são breves. O contrário do que ocorre no CDC, promulgado em 1990, em que, ao longo das várias modificações, vão ficando cada vez mais detalhados os trechos relativos a direitos e deveres e as definições só aparecem no início do texto de forma relativamente sucinta.

Uma outra observação a respeito de léxico se refere à presença dos *hápax legomena*, *tokens* de ocorrência única que aparecem nos códigos. Ambos os textos normativos apresentaram um número altíssimos deles: do total de *types* verbais são *hápax legomena* 53,85% no CA, e 48,13% no CDC. Isso pode indicar uma pequena queda na riqueza lexical, maior no texto mais antigo e menor naquele mais recente. Observando-se de perto tais verbos, a impressão que se tem é que, no CA, há um equilíbrio entre aqueles que tratam do tema das águas e aqueles que se referem ao próprio texto legislativo, como *promulgar*, *propor*, mas também *poluir*, enquanto no CDC há um número expressivo de verbos que abordam o tema do consumo, como *orçar*, *financiar*, *intimar*, entre muitos outros. Não foi realizada uma classificação completa que possa definir tal questão, pois outra pesquisa em curso está analisando o aspecto quantitativo na perspectiva da microdiacronia de um único código, de tamanho muito maior do que aqueles aqui analisados, o que poderá dar respostas mais precisas e com maior embase estatístico.

Um recorte da pesquisa foi verificar as ocorrências do verbo *ser*. Foram devidamente descartados os casos em que *ser* era utilizado como auxiliar, como na formação da passiva, e foi verificado que, em todas as ocorrências encontradas, *ser* era utilizado como verbo de ligação. Foi em seguida analisado a qual classe morfológica a palavra que se liga ao verbo na função de predicativo pertence. A Tabela 5 ilustra o que foi encontrado.

Tabela 5 – Predicativos utilizados com o verbo *ser* no CA e CDC em número absoluto e porcentagens

Categoria do predicativo	CA		CDC	
	N. ocorrências	% de uso	N. ocorrências	% de uso
adjetivo	47	60	15	65
substantivo	23	30	7	30
pronome	6	8	1	4
total	76	100	23	100

Fonte: Elaboração própria

Não foram percebidas grandes modificações do ponto de vista do tipo de predicativo, com um uso maior de substantivos, seguidos por adjetivos e, por fim, pelos pronomes. Exemplificamos abaixo com alguns trechos extraídos das leis:

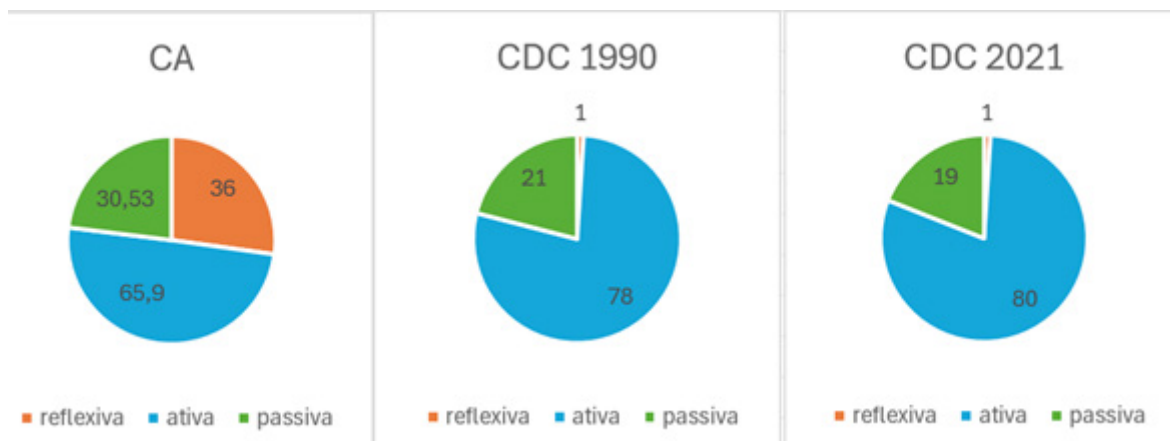
- (7) As concessões ou autorizações para derivação que se destinem a produção de energia hidro-elétrica **serão atribuições** aos Estados, na forma e com as limitações estabelecidas nos arts. 192, 193 e 194 (Brasil, 1934, cap. VII, art. 63, grifo nosso).
- (8) O Departamento Nacional de Defesa do Consumidor, da Secretaria Nacional de Direito Econômico (MJ), ou órgão federal que venha substituí-lo, **é organismo** de coordenação da política do Sistema Nacional de Defesa do Consumidor (Brasil, 1990, t. IV, art. 106, grifo nosso).
- (9) São particulares as nascentes e todas as águas situadas em terrenos que também **o sejam**, quando as mesmas não estiverem classificadas entre as águas comuns de todos, as águas públicas ou as águas comuns (Brasil, 1934, cap. III, art. 8, grifo nosso).

É interessante apontar que, justamente por constarem em trechos em que é definido o âmbito da aplicação da lei, os predicativos são lexicalmente constituídos por termos que remetem ao assunto abordado na norma. De fato, no CA os adjetivos mais utilizados são *público/os* (6 ocorrências no singular e 5 no plural), *navegável/navegáveis* (2 ocorrências no singular e 1 no plural), *flutuável/flutuáveis* (2 ocorrências no singular e 1 no plural) *particular* (4 ocorrências) e *comuns* (3 ocorrências). No CDC os adjetivos mais encontrados foram *impróprio*, *responsável/responsáveis*, *enganosa* e *competente* (2 ocorrências cada). Entre os substantivos mais utilizados tem-se, no CA, *águas* (3 ocorrências) e *requerente* (2 ocorrências) e, no CDC, *pessoa* (2 ocorrências). Já nos casos em que o predicativo é constituído por um pronome, houve uma mudança no tipo de pronome utilizado: enquanto no CA foram encontrados os clíticos *o* (3 ocorrências) e *lo* (3 ocorrências), no CDC foi localizado o demonstrativo *aquela* (1 ocorrência). A análise do teste qui-quadrado ($\chi^2 = 9,02$, $p = 0,011$, $gl = 2$) indica uma associação estatisticamente significativa entre as variáveis, com uma força de associação moderada (V de Cramér = 0,279). Esse resultado sugere que a mudança descrita não é devida ao acaso.

Uma possível explicação para a mudança do predicativo com pronome no texto mais recente, o CDC, pode ser atribuída à reorganização do quadro dos pronomes pessoais retratada por Castilho (2010), em que o pronome de terceira pessoa, prevalentemente utilizado nos textos do *corpus* aqui analisado, foi gradualmente desaparecendo em decorrência de um processo de redução fonológica, sendo substituído pelo demonstrativo.

Prosseguindo com os resultados das análises das tabelas Excel, são apresentadas, em seguida, as distinções em relação à voz verbal utilizada nos dois textos. Nesse caso, percebeu-se algumas pequenas, mas significativas mudanças, que são ilustradas na Imagem 2:

Imagem 2 – Comparação em gráficos de pizza das vozes verbais utilizadas no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem



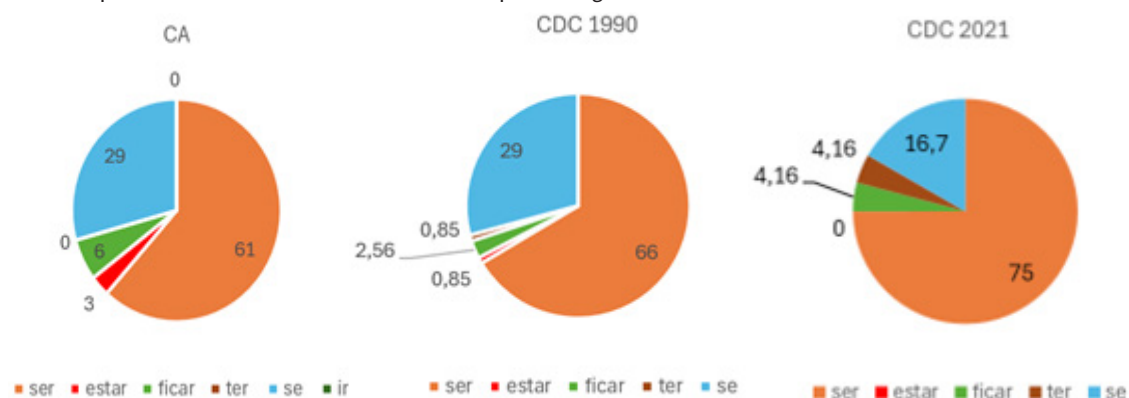
Fonte: Elaboração própria

Uma nova análise do teste qui-quadrado ($\chi^2 = 29,27$, $p = 0,000$, $gl = 4$) indicou uma associação estatisticamente significativa entre as variáveis, com uma força de associação moderada (V de Cramér = 0,091).

Lembrando que o CA foi promulgado em 1934 e, como dito, está em vigor até a data atual com pouquíssimas modificações, pode-se notar que há uma clara tendência no tempo à diminuição e quase extinção do uso da voz reflexiva e, em menor medida daquela passiva, ao observar-se os verbos extraídos dos dois períodos mais significativos do CDC em termos de modificações. Mais uma vez, não se pode afirmar que o léxico utilizado não tenha influenciado nas formas, e somente uma análise quantitativa maior poderá responder a tal questão.

Foram avaliados também os mecanismos de formação de passiva, para verificar se haveria algum tipo de mudança. A Imagem 3 ilustra os resultados.

Imagem 3 – Comparação em gráficos de pizza dos mecanismos de formação de passiva utilizados no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem



Fonte: Elaboração própria

O teste qui-quadrado ($\chi^2 = 48,61$, $p = 0,000$, $gl = 10$) indicou mais uma associação estatisticamente significativa entre as variáveis, com uma força de associação moderada (V de Cramér = 0,252).

As formas de apassivação utilizadas nos textos são auxiliarizações com os verbos *ser*, *estar*, *ficar*, *ter*, a partícula *se* apassivadora e, no CA, um único caso do verbo *ir*.

Vejamos alguns exemplos abaixo:

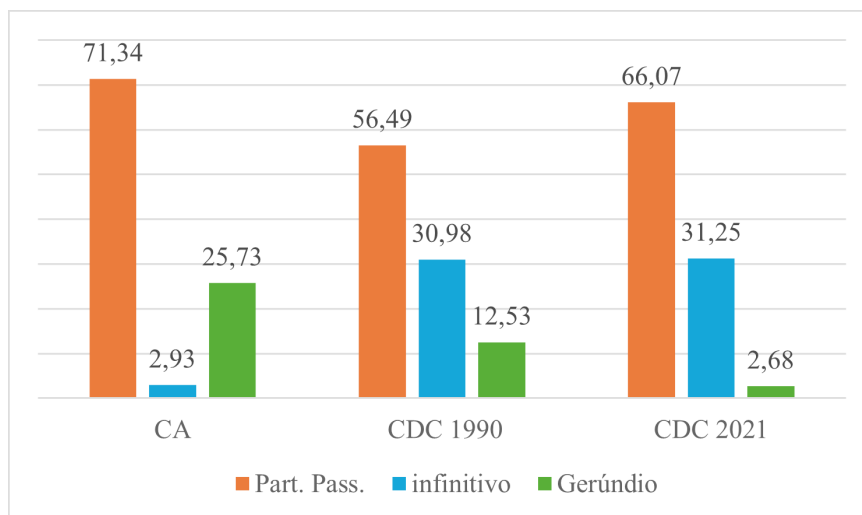
- (10) Quando a “avulsão” for de coisa não susceptível de aderência natural, **será regulada** pelos princípios de direito que regem a invenção (Brasil, 1934, cap. V, art. 21, grifo nosso).
- (11) Essa competência não exclui a dos Estados para legislarem subsidiariamente sobre a navegação ou flutuação dos rios, canais e lagos de seu território, desde que não **estejam compreendidos** nos números I e II do artigo 40 (Brasil, 1934, cap. VII, art. 61, par. ún., grifo nosso).
- (12) Nos contratos de adesão, o fornecedor [...] **fica obrigado** a entregar ao consumidor cópia do contrato, após a sua conclusão (Brasil, 1990, cap. VI-A, art. 54-G, § 2º, grifo nosso).⁹
- (13) Nos contratos do sistema de consórcio de produtos duráveis, a compensação ou a restituição das parcelas quitadas, na forma deste artigo, **terá descontada**, [...] os prejuízos que o desistente ou inadimplente causar ao grupo (Brasil, 1990, cap. VI, sec. II, art. 53, § 2º, grifo nosso).
- (14) Se o aqueduto tiver de atravessar estradas, caminhos e vias públicas, sua construção fica sujeita aos regulamentos em vigor, no sentido de não **se prejudicar** o trânsito (Brasil, 1934, t. VII, cap. ún., art. 122, grifo nosso).
- (15) Resolve decretar o seguinte Código de Águas, cuja execução compete ao Ministério da Agricultura e que **vai assinado** pelos ministros de Estado (Brasil, 1934, grifo nosso).

⁹ Incluído pela Lei nº 14.181, de 2021

Além do óbvio desuso do verbo *ir* como forma apassivadora, pode-se notar o aumento do verbo *ser* e uma diminuição da partícula *se*. Acreditamos que também nesse caso a redução no uso da forma *se* como traço de apassivação seja devida à mudança no quadro dos reflexivos indicado por Castilho (2010), pois o autor traça um panorama diacrônico do português brasileiro para mostrar como a interpretação passiva do reflexivo *se* permanece somente no estilo formal. Uma outra possível explicação, a ser melhor verificada, seria referente a mudanças nas normas de redação legislativa, com uma tendência ao uso de estruturas de mais fácil compreensão (Brasil, 2006).

As análises se debruçaram em outras classificações: não foram encontradas diferenças significativas nos distintos períodos quanto às categorias de pessoa e número, utilizando os textos sempre a terceira pessoa do singular e do plural. Quanto aos modos e tempos verbais, notam-se algumas mudanças significativas em seu uso. Em relação às formas nominais dos verbos, apresentamos os gráficos da Imagem 4.

Imagem 4 – Comparação em gráfico de barra da utilização das formas nominais dos verbos utilizados no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem



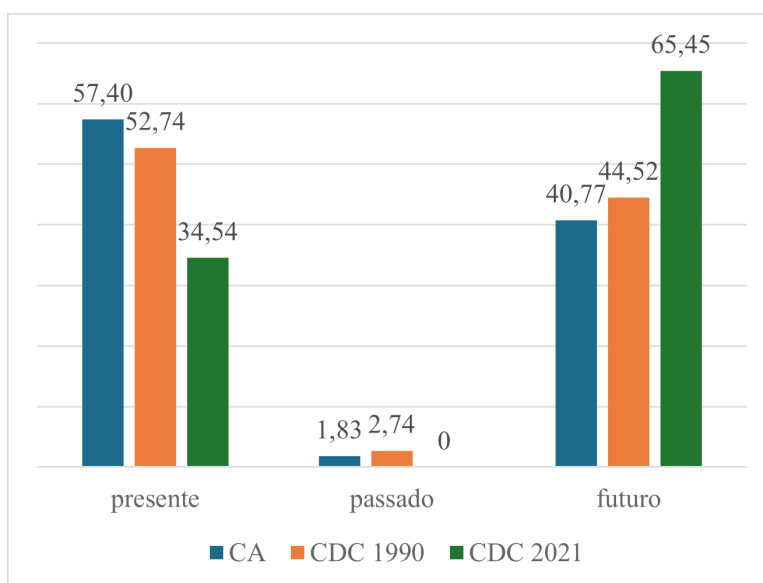
Fonte: Elaboração própria

Também nesse caso o teste do qui-quadrado ($\chi^2 = 114,7$, $p = 0,000$, $gl = 4$) indicou uma associação estatisticamente significativa entre as variáveis com uma força de associação moderada (V de Cramér = 0,259).

É possível perceber, nos períodos analisados, um grande aumento do uso do infinitivo, especialmente em detrimento do gerúndio. Tal dado é muito interessante, pois possibilita um recorte sobre uma mudança no tipo de orações subordinadas reduzidas nos dois textos. Há, no caso, uma dupla observação: do CA à versão original do CDC vê-se um enorme aumento no uso de infinitivas e diminuição das gerundivas. Já entre as duas versões do CDC, é observável uma estabilidade entre as infinitivas e uma diminuição expressiva das gerundivas. Quanto ao amplo uso do particípio passado, muitas vezes empregado como particípio adjetival, não foi possível identificar uma tendência clara, pois houve uma diminuição muito grande de 1934 para 1990, e depois um novo aumento nos dados de 2021. Não foi nosso objetivo nos aprofundar no assunto, mas certamente é algo a ser levado em conta em trabalhos futuros.

Em relação aos tempos verbais de indicativo e de subjuntivo, as análises agruparam os dados em macro categorias, visto que a maioria das ocorrências se dividiu entre presente e futuro nos dois modos. No indicativo passado, portanto, ficaram agrupados todos os casos de pretérito; e, no indicativo futuro, o futuro do presente e o futuro do pretérito. Já no subjuntivo o pretérito imperfeito foi categorizado como passado. As Imagens 5 e 6 mostram graficamente, em termos de porcentagem, as ocorrências encontradas nos códigos analisados.

Imagem 5 – Comparação em gráfico de barra da utilização dos tempos dos verbos utilizados no modo indicativo no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem



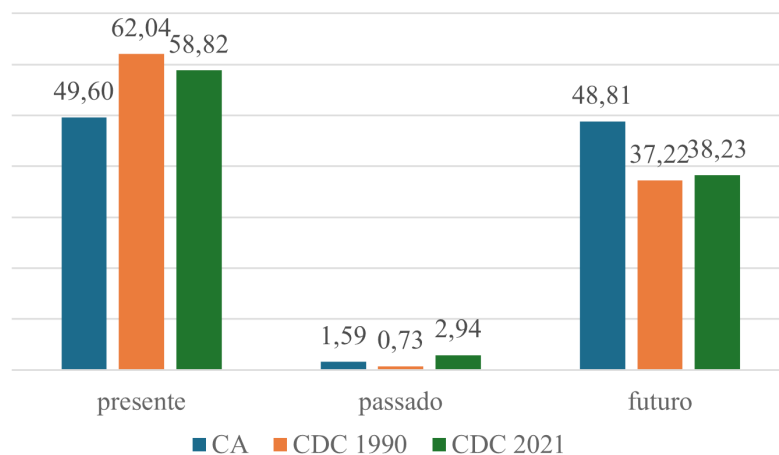
Fonte: Elaboração própria

Mais uma vez, o teste qui-quadrado ($\chi^2=13,88$, $p=0,008$, $gl=4$) indicou uma associação estatisticamente significativa entre as variáveis com uma força de associação moderada (V de Cramér = 0,09).

O gráfico da Imagem acima ilustra claramente uma diminuição no uso do presente do indicativo ao longo do tempo em favor do aumento do tempo futuro.¹⁰ Pode haver uma explicação para um incremento tão grande nas ocorrências do futuro em detrimento do presente no trecho do CDC de 2021: boa parte da modificação se refere a partes em que são citadas as medidas de prevenção e de tratamento do superendividamento (Capítulo VI-A) e isso requer do redator legislativo um amplo uso desse tempo para prever situações de aplicação da norma. Isso pode ter enviesado a amostra, apesar de ver-se que a tendência de utilização do futuro permanece.

¹⁰ Essa descoberta é especialmente importante porque um dos parâmetros da análise multidimensional que foi implementada no *corpus* inteiro apontou que os textos do *corpus* são direcionados ao futuro, inclusive, mas não somente, pelo amplo emprego dos tempos presente e futuro do indicativo e do subjuntivo (veja-se Marques et al., 2025).

Imagem 6 – Comparação em gráfico de barra da utilização dos tempos dos verbos utilizados no modo subjuntivo no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem

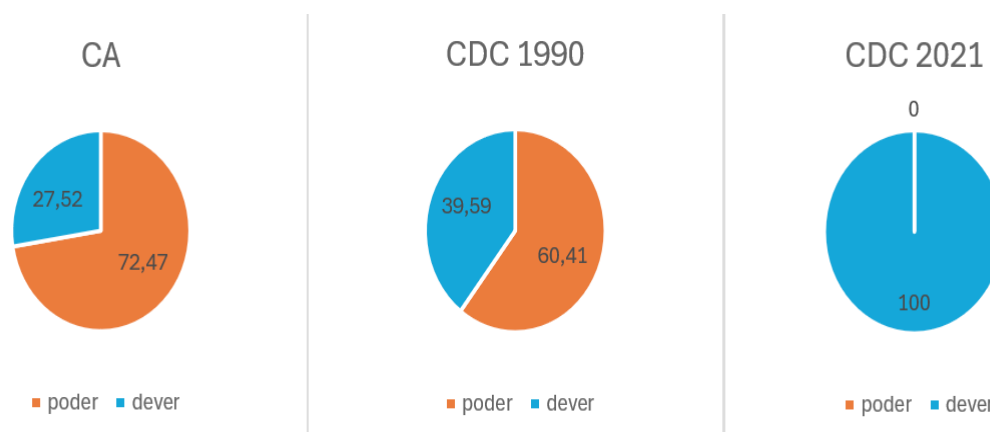


Fonte: Elaboração própria

Já no caso do subjuntivo, não foi possível visualizar uma tendência tão clara, com aumento e seguinte pequena diminuição do presente. De fato, nesse caso, o teste do qui-quadrado ($X^2 = 6,72$, $p = 0,152$, $gl = 4$) indicou que não há uma associação estatisticamente significativa entre as variáveis, sendo o valor crítico de 9,488 (V de Cramér = 0,089). Na referência ao futuro parece ter ocorrido uma diminuição na utilização do futuro e, em seguida, estabilidade. Como se trata de orações subordinadas, a análise deveria levar em conta fatores qualitativos mais refinados que não foram aqui implementados, mas que poderão ser analisados em outros trabalhos.

Entre as análises, foram verificados quais verbos modais foram utilizados nos dois códigos. A Imagem 7 abaixo ilustra os verbos encontrados e suas proporções, em porcentagem, nos três períodos.

Imagem 7 – Comparação em gráficos de pizza dos verbos modais utilizados no CA e no CDC nos períodos de 1990 e 2021, em termos de porcentagem



Fonte: Elaboração própria

O teste qui-quadrado ($\chi^2 = 22,94$, $p = 0,000$, $gl = 2$) indicou uma associação estatisticamente significativa entre as variáveis, com uma força de associação moderada (V de Cramér = 0,334).

Acreditamos que a totalidade do verbo *dever* no período de 2021 possa ter sido motivada pelo viés do trecho específico da lei que apontamos acima. De fato, a modificação inclui uma série de instruções sobre as modalidades de vendas à vista e a prazo como no Exemplo 16 abaixo:

- (16) Sem prejuízo do disposto no art. 37 deste Código, a oferta de crédito ao consumidor e a oferta de venda a prazo, ou a fatura mensal, conforme o caso, **devem indicar**, no mínimo, o custo efetivo total, o agente financiador e a soma total a pagar, com e sem financiamento. (Brasil, 1990, art. 54-B, § 3º, grifo nosso)

O mesmo viés legislativo apontado acima explicaria também o aumento do verbo *dever* em relação a *poder*. Uma análise qualitativa mais refinada e quantitativa no *corpus* inteiro, acompanhada de novos testes estatísticos, poderá responder a tal questão.

5 Conclusões

Este artigo apresentou uma análise em microdiacronia das mudanças das formas verbais e nominais dos verbos em dois Códigos extraídos do *corpus* LEX-BR-lus, devidamente anotado para tal finalidade. Os períodos analisados foram aqueles de 1934, para o Código de Águas, e de 1990 e 2021 para o Código do Consumidor. Os dados foram coletados manualmente e conferidos com as etiquetas do PALAVRAS (Bick, 2000, 2014), adotadas no *corpus*, e aquelas resultantes de uma análise multidimensional (Marques; Ferrari, 2023, Marques et al., 2025). Foi possível visualizar quantitativamente a variação lexical dos verbos, assim como dos tempos e modos verbais que compõem nas leis.

Após a tabulação dos dados e sua demonstração através de gráficos, pôde-se observar que, embora do ponto de vista dos lexemas verbais utilizados, os resultados são inconclusivos, pois acredita-se ser necessário um volume maior de dados, é possível entrever algumas tendências a respeito de modos e tempos verbais. Houve certa diminuição no uso do particípio passado, mas um considerável aumento do uso do infinitivo em detrimento do gerúndio. No modo indicativo, houve aumento do futuro em relação ao presente, enquanto no subjuntivo o aumento do presente e a diminuição do futuro requerem análises mais aprofundadas das subordinações. Foi constatado também um considerável aumento na utilização do verbo modal *poder* em detrimento de *dever*. Foi identificado, além disso, o quase desaparecimento da forma reflexiva e grande aumento da forma ativa. Entre as formas de passivação, o auxiliar *ser* foi o que mais cresceu, enquanto a partícula apassivadora se apresentou uma grande diminuição no uso. Entre os predicativos utilizados com o verbo *ser*, os adjetivos indicaram um ligeiro aumento.

Os testes estatísticos implementados (qui-quadrado, V de Cramér e cálculo do valor- p) permitem confirmar que, à exceção dos dados relativos ao subjuntivo, há uma significativa associação entre as variáveis com força de associação moderada, indicando não apenas que os resultados não são devidos ao acaso, mas também a necessidade de se ampliar o conjunto de dados para obter resultados mais precisos.

Apesar de serem iniciais, e possibilitarem refinamentos e desdobramentos, os resultados apresentados já mostraram algumas tendências muito interessantes de mudança na redação das normas. Pretende-se verificar, em trabalhos já iniciados e futuros, tais tendências, replicando as buscas em um número maior de textos e no *corpus* inteiro, com uma utilização mais ampla de ferramentas computacionais e testes estatísticos.

Como dito no início do artigo, trata-se ainda de observações preliminares em textos relativamente curtos. Todavia, as análises estão sendo refinadas e novos pontos de partida estão sendo traçados para uma varredura de alguns fenômenos em textos maiores, que possam dar um panorama mais amplo das mudanças da linguagem das leis ao longo do tempo no Brasil.

Alguns dos resultados estão sendo utilizados em novas pesquisas no mesmo *corpus* e poderão ser empregados por outros estudos em análises quantitativas e qualitativas em diversos textos do gênero legislativo.

Por fim, parece-nos que, apesar da ainda grande complexidade da linguagem das leis, especialmente para os leigos, o trabalho mostrou que há indícios de mudança e, quiçá, simplificação para uma sua maior legibilidade e clareza.

Agradecimentos

Agradecemos à Capes (bolsa n.º: 88887.626989/2021-00) e à Fapemig (bolsas de IC) pelo financiamento parcial da pesquisa.

Referências

ADLER, M. The Plain Language Movement. In: TIRESMA, P.; SOLAN, L. M. (Org.). *The Oxford Handbook of Language and Law*. New York: Oxford University Press, p. 67-83, 2012.

BARBERA, M.; ONESTI, C. Scheda progetto di ricerca n. 9. *Corpus Jus Jurium*. In: DIADORI, P. *Progetto JURA: la formazione dei docenti di lingua e traduzione in ambito giuridico italo*. Perugia: Guerra Edizioni, p. 349-351, 2009.

BERBER SARDINHA, T.; KAUFFMANN, C.; ACUNZO, C. M. Dimensions of Register Variation in Brazilian Portuguese. In: PINTO, M. V. (Ed.). *Multi-Dimensional Analysis: 25 Years on a Tribute to Douglas Biber*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014.

BERBER SARDINHA, T.; PINTO, M. V. (Ed.). *Multi-Dimensional Analysis: Research Methods and Current Issues*. Londres: Bloomsbury Academic, 2019.

BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge University Press, 1988.

BIBER, D. *Dimensions of Register Variation: A Cross-linguistic Perspective*. Cambridge: Cambridge University Press, 1995.

BIBER, D. Representativeness in *Corpus Design*. *Literary and Linguistic Computing*, Oxford, v. 8, n. 4, p. 243-257, 1993. DOI: <https://doi.org/10.1093/lc/8.4.243>.

BICK, E. PALAVRAS, a Constraint Grammar-based Parsing System for Portuguese. In: BERBER SARDINHA, T.; FERREIRA, T. (Ed.). *Working with Portuguese Corpora*. London: Bloomsbury, 2014. p. 279-302.

- BICK, E. The Parsing System “PALAVRAS”: *Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. 2000. Tese (Doutorado em Linguística) – Department of Linguistics, Aarhus University, Aarhus, 2000.
- BIDERMAN, M. T. C. *A face quantitativa da linguagem: um dicionário de frequências do português*. ALFA: Revista de Linguística, São Paulo, v. 42, n. 1, 2001.
- BIEL, Ł. Corpora in Institutional Legal Translation: Small Steps and the Big Picture. In: RAMOS, F. P. Institutional Translation for International Governance. Enhancing Quality in Multilingual Legal Communication. London: Bloombury Academic, p. 25-26, 2018.
- BRASIL. Casa Civil. *Subchefia para Assuntos Jurídicos*. Legislação – Presidência da República. Disponível em: <https://www4.planalto.gov.br/legislacao/>. Acesso em: 30 jan. 2025.
- BRASIL. *Constituição da República Federativa do Brasil de 1988*. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 20 jan. 2025.
- BRASIL. *Decreto nº 24.643, de 10 de julho de 1934*. Institui o Código de Águas. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto/d24643compilado.htm. Acesso em: 20 jan. 2025.
- BRASIL. Lei nº 14.181, de 1º de julho de 2021. Altera a Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor), e a Lei nº 10.741, de 1º de outubro de 2003 (Estatuto do Idoso), para aperfeiçoar a disciplina do crédito ao consumidor e dispor sobre a prevenção e o tratamento do superendividamento. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14181.htm. Acesso em: 20 jan. 2025.
- BRASIL. *Lei nº 7.347, de 24 de julho de 1985*. Disciplina a ação civil pública de responsabilidade por danos causados ao meio-ambiente, ao consumidor, a bens e direitos de valor artístico, estético, histórico, turístico e paisagístico (VETADO) e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l7347orig.htm. Acesso em: 20 jan. 2025.
- BRASIL. *Lei nº 8.078, de 11 de setembro de 1990*. Dispõe sobre a proteção do consumidor e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 20 jan. 2025.
- BRASIL. Presidência da República. Casa Civil. *Manual de redação da Presidência da República / Casa Civil, Subchefia de Assuntos Jurídicos*. Coordenação de Gilmar Ferreira Mendes, Nestor José Forster Júnior [et al.]. 3. ed. rev. atual. e ampl. Brasília: Presidência da República, 2018.
- BRASIL. Senado Federal. *Consultoria Legislativa*. Manual de Redação Parlamentar e Legislativa. Brasília: Senado Federal, Consultoria Legislativa, 2006.
- BREZINA, V. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press, 2018.
- CANAVESE, P. LEX.CH.IT: A *Corpus* for Micro-Diachronic Linguistic Investigations of Swiss Normative Acts in Italian. *Comparative Legilinguistics*, v. 40, p. 44-65, 2019. Disponível em: <https://doi.org/10.14746/cl.2019.40.3>.
- CAO, Y.; XIAO, R. A Multi-dimensional Contrastive Study of English Abstracts by Native and Non-Native Writers. *Corpora*, v. 8, n. 2, p. 209-234, 2013. DOI: <https://doi.org/10.3366/cor.2013.0041>.
- CASTILHO, Ataliba T. de. *Nova gramática do Português Brasileiro*. São Paulo: Editora Contexto, 2010.

- CEGALLA, D. P. *Novíssima Gramática da Língua Portuguesa*. 48. ed. São Paulo: Companhia Editora Nacional, 2011.
- CETESB – COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. *Histórico da legislação hídrica no Brasil*. Águas Interiores, [2023]. Disponível em: <https://cetesb.sp.gov.br/aguas-interiores/tiipos-de-agua/historico-da-legislacao-hidrica-no-Brasil/>. Acesso em: 13 dez. 2024.
- CORTELAZZO, M. L'italiano giuridico sta davvero cambiando? In: POZZO, B.; BAMBI, F. (Org.). *L'italiano giuridico che cambia*. Firenze: Accademia della Crusca, 2012. p. 179-183.
- CORTELAZZO, M. Lingua e diritto in italia. Il punto di vista dei linguisti. In: SCHENA, L. (Ed.). *La lingua del diritto. Difficoltà traduttive. Applicazioni didattiche. Atti del primo Convegno Internazionale*, Milano, 5-6 ottobre. Roma: Cisu (Centro d'Informazione e Stampa Universitaria), 1997. p. 35-50.
- COULTHARD, M.; JOHNSON, A. *An Introduction to Forensic Linguistics: Language in Evidence*. New York: Routledge, 2007.
- COULTHARD, M.; JOHNSON, A. *The Routledge Handbook of Forensic Linguistics*. Abingdon: Routledge, 2010.
- CRAMÉR, H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.
- DAHLMAN, R. C. *Specialità del linguaggio giuridico italiano*. Stockholm University, 2006. Disponível em: https://www.researchgate.net/publication/27818945_Specialita_del_linguaggio_giuridico_italiano. Acesso em: 6 set. 2020.
- DAVIES, M.; FERREIRA, A. *Corpus do Português: 45 Million Words, 1300s-1900s*. 2006. Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 15 set. 2022.
- FERRARI, L. A.; CUNHA, E. L. T. P. Reflexões metodológicas sobre datasets e Linguística de Corpus: uma análise preliminar de dados legislativos. *Domínios de Linguagem*, Uberlândia, v. 16, n. 4, p. 1571-1607, 2022. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/64146>. Acesso em: 23 set. 2022.
- FERRARI, L. A.; MARQUES, C. G. F. O LEX-BR-lus: arquitetura e decisões na compilação de um corpus representativo das leis federais brasileiras. *ANTARES*, v. 14, n. 34, 2022. Disponível em: <http://www.uces.br/etc/revistas/index.php/antares/article/view/11150/5328>. Acesso em: 19 dez. 2022.
- FERRAZ JUNIOR, T. S. *O arcaísmo na formação jurídica brasileira*. Revista USP, v. 1, p. 28-36, 2014.
- FISHER, R. A. *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd, 1925.
- GIAMPIERI, P. A Diachronic Analysis of Latinisms in the Decisions of the UK Supreme Court. *Int. J Semiot Law*, 2024. DOI: <https://doi.org/10.1007/s11196-024-10232-1>.
- GIAMPIERI, P. Online Parallel and Comparable Corpora for Legal Translations. *Altre modernità*, n. 20, p. 237-252, 2018. DOI: <https://doi.org/10.13130/2035-7680/10842>. Disponível em: https://www.researchgate.net/publication/329365608_Online_Parallel_and_Comparable_Corpora_for_Legal_Translations. Acesso em: 6 set. 2020.
- GOŹDŹ-ROSKOWSKI, S. Corpus Linguistics in Legal Discourse. *International Journal for the Semiotics of Law*, n. 34, p. 1515–1540, 2021. DOI: <https://doi.org/10.1007/s11196-021-09860-8>.
- GOŹDŹ-ROSKOWSKI, S. Legal Language. In: CHAPPELLE, Carol A. (Org.). *The Encyclopedia of Applied Linguistics*. Chichester?: John Wiley e Sons, 2012. p. 3281–3287.

GOŹDŹ-ROSZKOWSKI, S. *Patterns of Linguistic Variation in American Legal English: A Corpus-Based Study*. Frankfurt am Main: Peter Lang, 2011.

GOŹDŹ-ROSZKOWSKI, S.; PONTRANDOLFO, G. Legal Phraseology Today: *Corpus*-based Applications Across Legal Languages and Genres. *International Journal of Specialized Communication*, 37(3-4):130-138 2015. Disponível em: https://www.academia.edu/18714805/Legal_Phraseology_Today_Corpus_based_Applications_Across_Legal_Languages_and_Genres. Acesso em: 6 set. 2020.

GUGLINSKI, V. *Breve histórico do Direito do Consumidor e origens do Código de Defesa do Consumidor*. Meu Site Jurídico, 8 maio 2019. Disponível em: <https://meusitejuridico.editorajuspodivm.com.br/2019/05/08/breve-historico-direito-consumidor-e-origens-codigo-de-defesa-consumidor/>. Acesso em: 15 dez. 2024.

HARDIE, A. Modest XML for Corpora: not a standard, but a suggestion. *ICAME Journal*, v. 38, n. 1, p. 73–103, 2014. DOI: <https://doi.org/10.2478/icame-2014-0004>. Acesso em: 20 jul. 2021.

JUSBRAZIL. *Portal Jurídico JusBrasil*. Disponível em: <https://www.jusBrasil.com.br/>. Acesso em: jun. 2023.

MARQUES, C. G. de F. *Análise multidimensional dos textos legais federais brasileiros*. 2023. Dissertação (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2023.

MARQUES, C. G. de F; FERRARI, L. A. Características gerais das leis federais brasileiras: um estudo exploratório sob o enfoque da Análise Multidimensional. In: *SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL)*, 14., 2023, Belo Horizonte/MG. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 308-317. DOI: <https://doi.org/10.5753/stil.2023.233485>.

MARQUES, F. P. A.; FERRARI, A. L.; KAUFFMANN, C. H. Challenges in a Multidimensional Analysis of a Brazilian Portuguese *Corpus* of Federal Statutory Laws. *Revista Brasileira de Linguística Aplicada*, v. 25, n. 1, p. 1-38, 2025. DOI: <https://doi.org/10.1590/1984-6398202549849>. Disponível em: <https://www.scielo.br/j/rbla/a/s4PcfM453T75kbvwwJfTQDd/?lang=en#>. Acesso em: 11 dez. 2025.

ONESTI, C. *Methodology for Building a Text-Structure Oriented Legal Corpus*. *Comparative Legilinguistics*, v. 8, p. 37- 48, 2011. DOI: <https://doi.org/10.14746/cl.2011.08.04>.

PEARSON, K. On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it Can be Reasonably Supposed to Have Arisen from Random Sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 50, n. 302, p. 157-175, 1900. DOI: <https://doi.org/10.1080/14786440009463897>.

PONTRANDOLFO, G. Legal Corpora: an Overview. *Rivista Internazionale di Tecnica della Traduzione*, p. 121-136, *Trieste*, 2012. Disponível em: <https://www.openstarts.units.it/bitstream/10077/9783/1/12Pontrandolfo.pdf>. Acesso em: 6 set. 2020.

RICHARD, I. Is Legal Lexis a Characteristic of Legal Language? *Lexis*, n. 11, p. 1-14, 2018. DOI: <https://doi.org/10.4000/lexis.1173>. Disponível em: https://www.researchgate.net/publication/324949333_Is_legal_lexis_a_characteristic_of_legal_language. Acesso em: 6 set. 2020.

RODRIGUES, M. C. C. *Linguagem, discurso e direito: algumas questões de linguística jurídica*. *Revista Do Ministério Público*, n. 111, p. 5-37, jul.-set. 2007.

RODRIGUES, R.; LEMOS-COUTO, M.; LEME COELHO, F.; SOUZA DE MIRANDA JUNIOR, I.; VALE, O. Bases lexicais verbais do português Brasileiro. *Domínios de Linguagem*, Uberlândia, v. 16, n. 4, p. 1489-

1516, 2022. DOI: <https://doi.org/10.14393/DL52-v16n4a2022-9>. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/63470>. Acesso em: 2 jan. 2025.

SINCLAIR, J. *Trust the Text. Language, Corpus, and Discourse*. London: Routledge, 2004.

TIERSMA, P. M. *Legal Language*. Chicago: The University of Chicago Press, 2000.

VISCONTI, J. Prestiti e calchi: dove va la lingua giuridica italiana. In: POZZO, B.; BAMBI, F. (Org.). *L'italiano giuridico che cambia*. Firenze: Accademia della Crusca, p. 185-193, 2012.

VISCONTI, J. Interferenze sull'italiano giuridico nei processi di traduzione e di trasposizione dalla lingua inglese. In: SENATO DELLA REPUBBLICA (Org.). *Atti della Giornata di studio Il linguaggio giuridico nell'Europa delle pluralità. Lingua italiana e percorsi di produzione e circolazione del diritto dell'Unione europea*. Roma: Senato della Repubblica, p. 71-81, 2017.

WILLIAMS, C. Court judgments and plain language: Some recent developments in Canada and the United Kingdom. In: MACI, S.; SALA, M.; SPINZI, C. (Ed.). *Communicating English in Specialised Domains: A Festschrift for Maurizio Gotti*. Newcastle upon Tyne: Cambridge Scholars Publishing, p. 106-125, 2020.

WILLIAMS, C. *Legal English and Plain Language: An Introduction*. *ESP Across Culture*, v. 1, p. 111-124, 2004.

WILLIAMS, C. *The Impact of Plain Language on Legal English in the United Kingdom*. London: Routledge, 2023.