

Características gerais das leis federais brasileiras: um estudo exploratório sob o enfoque da Análise Multidimensional

Carolina Godoi de Faria Marques, Lúcia de Almeida Ferrari

Faculdade de Letras - Universidade Federal de Minas Gerais

carol.godoi@outlook.com.br, ferrari.lu@gmail.com

Abstract. *Legal language is a rich and multifaceted specialized language used in various contexts, such as in courts, laws, and law textbooks [Goźdź-Roszkowski 2012; Carapinha 2018]. Despite its richness, there is a paucity of studies describing the linguistic features of this language in Brazilian Portuguese. Aiming to contribute to this field, we investigated its use in Brazilian federal statutory legal texts analyzing their linguistic variation and verifying if they are a register [Biber and Conrad 2009]. To do so, we conducted, supported by Corpus Linguistics, a Multidimensional Additive Analysis [Biber 1988] of the LEX-BR-Ius corpus [Ferrari and Marques in compilation] and confirmed our hypothesis.*

Resumo. *A linguagem jurídica é uma linguagem especializada rica e multifacetada utilizada nos mais diversos contextos, tais como em tribunais, leis e livros didáticos de direito [Goźdź-Roszkowski 2012; Carapinha 2018]. Apesar da sua riqueza, há poucos estudos que descrevem as características linguísticas dessa linguagem no português brasileiro. No intuito de contribuir com tal campo escolhemos investigar os textos legais federais brasileiros analisando sua variação linguística e verificando se eles são um registro [Biber e Conrad 2009]. Para tanto, realizamos, guiados pela Linguística de Corpus, uma Análise Multidimensional Aditiva [Biber 1988] do corpus LEX-BR-Ius [Ferrari e Marques em preparação e 2022] e confirmamos nossa hipótese.*

1. Introdução

Este trabalho apresenta os resultados de uma pesquisa de mestrado que realizou uma Análise Multidimensional aditiva [Biber 1988] utilizando os dados do *LEX-BR-Ius* [Ferrari e Marques em preparação e 2022] e se insere no projeto homônimo que busca descrever linguisticamente os textos legais federais brasileiros. Iremos aqui detalhar as particularidades de tais textos sob o enfoque da Análise Multidimensional [Biber 1988], uma abordagem empírico-metodológica que se vale de métodos quantitativos para identificar e medir as características linguísticas de um dado corpus, calculando a variação nos textos que o compõem e interpretando seus resultados de forma qualitativa. Graças a tal metodologia, foi possível obter um panorama da linguagem utilizada nesses textos e verificar nossa hipótese de que os textos legais são um registro segundo a perspectiva de Biber e Conrad (2009).

A linguagem jurídica se insere no rol das linguagens especializadas por apresentar léxico e gramática específicos. É importante lembrar que sob tal especificação há uma variedade muito grande de situações, e, portanto, linguagens diversas, que vão desde textos escritos, como as normas jurídicas e as várias etapas processuais, a interações orais,

como os depoimentos e oitivas entre muitos outros [Tiersma 1999; Coulthard e Johnson 2007, 2010; Goźdż-Roszkowski 2012; Carapinha 2018; Lorz 2019]. Delimitar e descrever a linguagem jurídica é, portanto, tarefa árdua e longa que requer recortes específicos. Acreditamos estar contribuindo para tal através dessa pesquisa que se configura como uma etapa preliminar na busca de uma descrição detalhada da língua empregada nas normas legais brasileiras.

O trabalho de Berber Sardinha, Kauffmann e Acunzo (2014) descreveu, valendo-se da Análise Multidimensional do Corpus Brasileiro de Variação e Registro (CBVR) [Berber Sardinha, Kauffmann e Acunzo 2014], as características linguísticas de uma grande variedade de registros do português brasileiro (doravante PB), traçando desta forma um retrato da variação linguística no PB. O trabalho que apresentamos somou o *LEX-BR-Ius* às dimensões identificadas por Berber Sardinha, Kauffmann e Acunzo (2014), comparando-o com o CBVR e buscando evidenciar as diferenças entre eles e identificar as características próprias do nosso corpus.

Como dito acima, a linguagem jurídica representa um universo multifacetado que merece ser analisado em cada uma de suas especificidades. Assim como uma poesia é muito diferente de um romance ou de um conto ou de um artigo de jornal, normas legais são tipos textuais muito distintos de julgamentos, petições, sentenças e assim por diante. Mostraremos, nesta contribuição, que não somente as normas legais são um registro específico, mas também que os critérios de seleção para a sua análise resultam em aspectos linguísticos específicos.

2. Referenciais teóricos e metodológicos

Um corpus é um conjunto de textos autênticos, coletados de acordo com critérios específicos de arquitetura para que seja representativo de uma língua ou de uma determinada variedade linguística, armazenado e submetido a tratamento computacional, possibilitando buscas automáticas ou semiautomáticas [Sinclair 2004; McEnery e Wilson 2001; Tognini-Bonelli 2001; Baker, Hardie e McEnery 2006].

Dentre os métodos de análise de corpus temos a Análise Multidimensional (AMD) [Biber 1988] utilizada para o estudo da variação linguística. A AMD permite tanto o estudo sincrônico quanto diacrônico de quaisquer línguas. Para tanto, esse método adota a perspectiva do registro, definido por Biber e Conrad (2009) como uma variedade da língua com traços situacionais, linguísticos e funcionais específicos, utilizada em contextos comunicativos próprios. Conforme Biber (1988), a AMD tem como pressuposto que a co-ocorrência de traços linguísticos em determinados padrões é fruto de fatores funcionais a eles inerentes. A identificação desses padrões em textos de determinado registro e comparação com outros registros possibilita sua caracterização. Para a descrição da sua variação, os registros são distribuídos segundo seus traços linguísticos em dimensões. Um registro tem várias dimensões intrínsecas, reveladas a partir da submissão dos dados de um corpus a uma série de operações estatísticas e seus resultados a uma interpretação funcional. [Biber 1988; Biber e Conrad 2009; Berber Sardinha, 2010]

O termo variação se refere aqui às diferenças entre a linguagem natural utilizada na comunicação, escrita e falada, de informantes diversos na mesma situação e contexto de uso ou pelo mesmo informante em diafásias distintas [Biber e Conrad 2009]. A AMD evidencia quais traços morfológicos e sintáticos do registro predominam em cada

dimensão selecionada pelo pesquisador possibilitando cruzar tais informações e traçando, de tal maneira, um quadro preciso das características linguísticas do registro em análise [Biber 1988].

Segundo Biber (1988) e Berber Sardinha (2013a), ao adotar a AMD como método para estudar a variação linguística devemos escolher entre realizar a Análise Multidimensional completa (*Full MD*) ou a Análise Multidimensional aditiva (Aplicação de dimensões). Embora ambas forneçam dados linguísticos significativos que permitem o estudo da variação linguística nos registros analisados, elas se diferenciam pela profundidade e finalidade da análise, assim como pela metodologia empregada, se tratando assim de abordagens complementares.

A primeira identifica as características linguísticas dos registros em análise e estabelece suas dimensões de variação, fornecendo uma descrição precisa e detalhada seja dos registros que das dimensões. Para tanto, várias etapas metodológicas quantitativas e qualitativas devem ser seguidas. Esse tipo de AMD, entretanto, é limitada em escopo, uma vez que abrange uma quantidade finita de registros e seu rigor metodológico, especialmente em relação às exigências na compilação do corpus a ser analisado, e a complexidade dos cálculos estatísticos requeridos muitas vezes representa um impedimento para a sua realização. [Biber 1988; Biber e Conrad 2009; Berber Sardinha 2013a]

Já a segunda incorpora os registros analisados a dimensões já identificadas por uma AMD completa, enriquecendo-a, pois aumenta seu escopo. Esses registros são então comparados com aqueles utilizados para identificar as dimensões às quais o corpus foi adicionado e o resultado é um panorama das características linguísticas e da variação dos registros estudados ao longo das dimensões. Para isso, o corpus utilizado deve ser compilado seguindo os mesmos parâmetros do corpus utilizado no estudo que identificou as dimensões adotadas. É importante ressaltar também que cabe ao pesquisador escolher em quais das dimensões adicionará seu corpus, podendo adicioná-lo a todas ou apenas algumas. A realização da AMD aditiva é considerada mais simples, flexível e rápida, pois comprehende menos etapas e procedimentos estatísticos, sendo eles menos complexos, quando comparada à primeira, e tem um rigor metodológico menor. [Biber 1988; Berber Sardinha 2013a]

3. Materiais e métodos

Para realizar a pesquisa utilizamos dados extraídos do corpus *LEX-BR-Ius*, um corpus sincrônico de normas legais federais brasileiras que apresenta, entre suas escolhas metodológicas de representatividade: (a) textos inteiros de maneira a garantir sua organização interna e conteúdo [Sinclair 2004; Biber 1993; Ferrari e Marques 2022; Marques 2023]; b) textos selecionados com base em sua frequência de uso no mundo legal [Barbera e Onesti 2009; Ferrari e Marques 2022; Marques 2023]; normas legais em vigor no momento de sua extração [Ferrari e Marques 2022; Marques 2023]. O corpus é dividido em seções segundo o tipo de norma, seguindo a classificação adotada pelo Planalto Brasileiro: Constituição, Códigos, Estatutos, Emendas à Constituição, Leis complementares e Leis ordinárias.

Com o objetivo de identificar os traços linguísticos característicos dos textos legais federais brasileiros e determinar se eles podem ser considerados um registro submetemos nosso corpus a uma Análise Multidimensional Aditiva com o auxílio dos softwares: Microsoft Excel e “IBM SPSS Statistics 23” [IBM 2015]. Esclarecemos que,

para atender às exigências da AMD aditiva o corpus foi etiquetado morfossintaticamente e lematizado com o o PALAVRAS [Bick 2000 e 2014], o mesmo etiquetador utilizado no estudo que identificou as dimensões às quais adicionamos o corpus. Foi necessário também fazer um recorte nos dados, sendo utilizadas apenas as seções: Constituição, Códigos e Estatutos, de forma a evitar possíveis vieses e permitir uma comparação a mais fidedigna possível dos nossos dados àqueles do estudo anterior.

A análise foi precedida de um levantamento preliminar das pesquisas sobre linguagem jurídica, com foco naquelas que compreendiam os textos legais. Para a realização da AMD aditiva em si, optamos por utilizar os dados da Análise Multidimensional do CBVR adicionando nosso corpus a todas as dimensões de variação do português brasileiro por ela identificadas, quais sejam: (1) *Oral versus literate discourse*, (2) *Argumentation*, (3) *Involved versus informational production*, (4) *Directive discourse*, (5) *Future versus past time orientation* e (6) *Reported discourse*. Para tanto, primeiramente individualizamos os traços linguísticos presentes em cada uma dessas dimensões a partir dos dados fornecidos em Berber Sardinha, Kauffmann e Acunzo (2014). Em seguida, submetemos nosso corpus anotado a um tratamento no pós-processador “PALAVRAS Tag count” [Berber Sardinha 2013b] que realiza a contagem automática das ocorrências dos traços em cada texto do corpus.

A fim de equilibrar os dados e evitar vieses, normalizamos por mil palavras as ocorrências dos traços contabilizadas anteriormente e calculamos seus Z-escores. Posteriormente, calculamos o escore de dimensão de cada texto que compõe o corpus ao longo de cada dimensão analisada a fim de obter a carga fatorial de cada um deles. Já para obter as cargas fatoriais do corpus em cada dimensão, de forma a observar em que ponto da escala de cada uma delas ele se localiza calculamos a média dos escores de dimensão. Por fim, adicionamos os dados obtidos à planilha das médias de dimensão do CBVR a nós fornecida pelos autores do estudo e comparamos nossos dados àqueles. Para complementar nossa análise, submetemos os dados aos testes estatísticos ANOVA e R^2 e calculamos o *p-value* para verificar o grau de variação abarcado pelo corpus em cada dimensão e se ele se diferencia dos registros do CBVR de forma estatisticamente relevante.

4. Resultados

Conforme apresentado na seção anterior, comparamos nossos dados com o CBVR em todas as dimensões do PB. Nossa corpus obteve pontuações únicas em todas elas, entretanto, as pontuações obtidas nas dimensões 2 (*Argumentation*), 4 (*Directive discourse*) e 6 (*Reported discourse*) giraram em torno de 0. Diante desses resultados concluímos que a argumentação (dimensão 2) não exerce um papel fundamental nos textos legais, não sendo eles marcados pela presença substantiva de traços argumentativos nem pela sua ausência. Da mesma forma, os textos não são marcados pela instrucionalidade nem pela sua falta (dimensão 4). Identificamos, no nosso corpus, a presença de alguns traços linguísticos geralmente associados a ordens, instruções ou orientações para a execução de determinadas tarefas, mas eles apresentam pontuações tanto positivas quanto negativas baixas. Por fim, também o discurso indireto (dimensão 6) não é característico no nosso corpus. Embora alguns de seus traços estejam presentes nos textos legais sua pontuação não é suficiente para caracterizá-los como marcados pelo discurso indireto nem pela sua ausência. Provavelmente, como os textos legais são

impositivos, criados com o objetivo de informar e descrever as normas que regem os sujeitos de direito da forma mais clara e detalhada possível para evitar que hajam interpretações diversas da pretendida, a argumentação e a instrucionalidade exercem um papel secundário no seu texto, sendo o discurso direto preferido àquele indireto.

Por esta razão, aprofundaremos a seguir somente as dimensões mais relevantes para a caracterização dos nossos dados, ou seja, aquelas nas quais houve uma maior pontuação, quais sejam: 1, 3 e 5 e reproduziremos apenas parte dos dados referentes aos registros do CBVR.

4.1. Dimensão 1: *Oral versus literate discourse*

Essa dimensão abrange as características do discurso oral e do discurso letrado: ao longo do polo positivo estão os registros típicos do discurso oral e daquele negativo aqueles típicos do discurso letrado.

A seguir reproduzimos o gráfico dessa dimensão com a adição de nosso corpus.

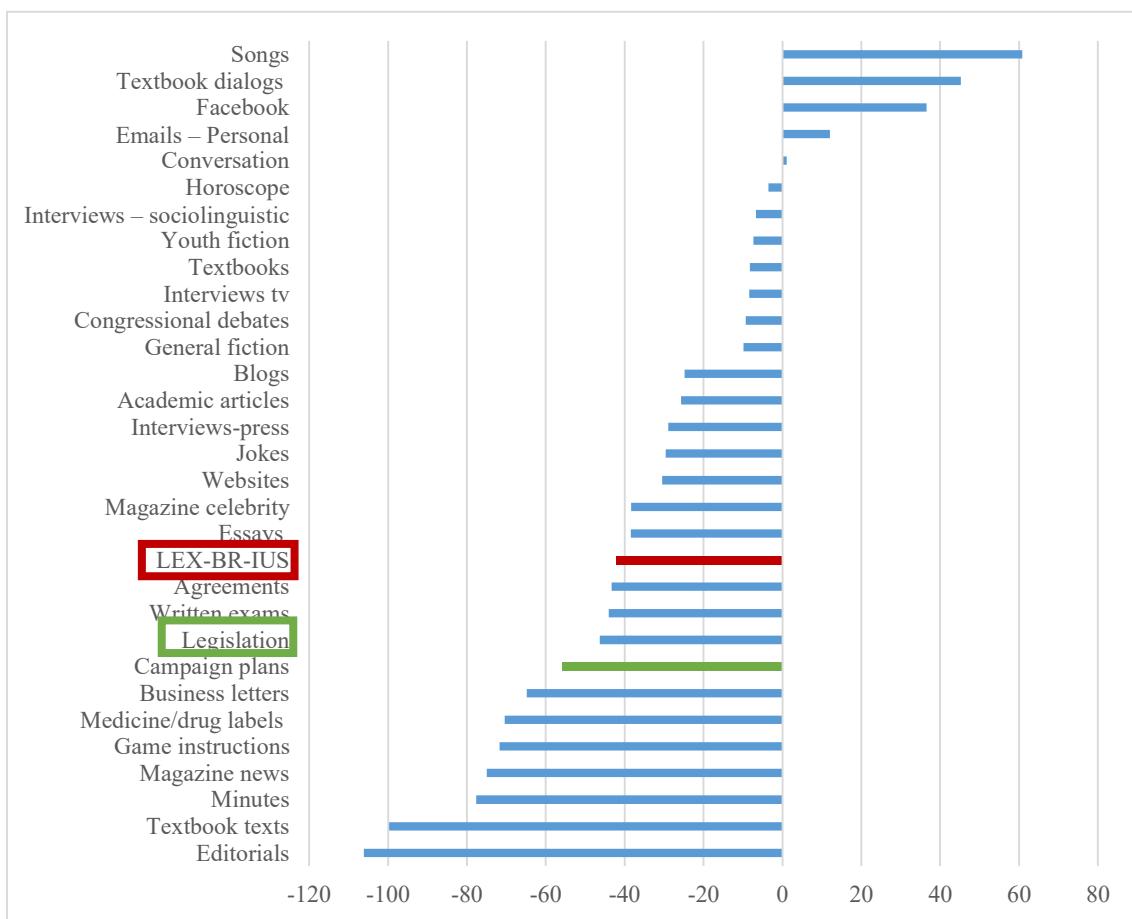


Figura 1. Dimensão 1

Tanto nosso corpus quanto a seção “*Legislation*” do CBVR se localizam no polo negativo, sendo que a seção “*Legislation*”, pontuou de forma mais negativa que o nosso corpus, o que significa que, apesar de ambos serem discursos letrados, os textos dessa seção do CBVR têm uma maior ocorrência de traços típicos desse discurso que aqueles

do *LEX-BR-Ius*. Os registros desse polo são marcados pela alta densidade informacional, discurso formal e especializado. Dentre os traços típicos desse polo destacamos a forte presença de sintagmas nominais, substantivos compostos e abstratos, adjetivos, artigos definidos, preposições, particípios passados e nominalizações na posição de sujeito. A seguir trazemos exemplos dessas características nos nossos textos:

(i) Art. 1º Esta Lei institui o Estatuto da Igualdade Racial, destinado a garantir à população negra a efetivação da igualdade de oportunidades, a defesa dos direitos étnicos individuais, coletivos e difusos e o combate à discriminação e às demais formas de intolerância étnica. [Brasil 2010]

Nos exemplos acima reproduzidos temos uma presença expressiva de vários dos traços típicos dessa dimensão. Destacamos a presença de diversos sintagmas nominais de vários tamanhos, como: Esta Lei; o Estatuto da Igualdade Racial. Ressaltamos também o uso de vários substantivos abstratos como: discriminação e adjetivos como: negra, étnicos, etc. Citamos ainda o uso de artigos definidos e preposições simples e compostas. Temos também a presença de particípios passados como: destinado. Todos esses traços contribuem para a especificação, restrição e detalhamento do conteúdo veiculado explicam a alta densidade informacional e complexidade das estruturas sintáticas encontradas no nosso corpus.

4.2 Dimensão 3: *Involved versus informational production*

Nessa dimensão os registros estão dispostos segundo o nível de envolvimento/interação dos participantes no registro no polo positivo e seu caráter informativo no polo negativo, conforme pode ser observado no gráfico abaixo.

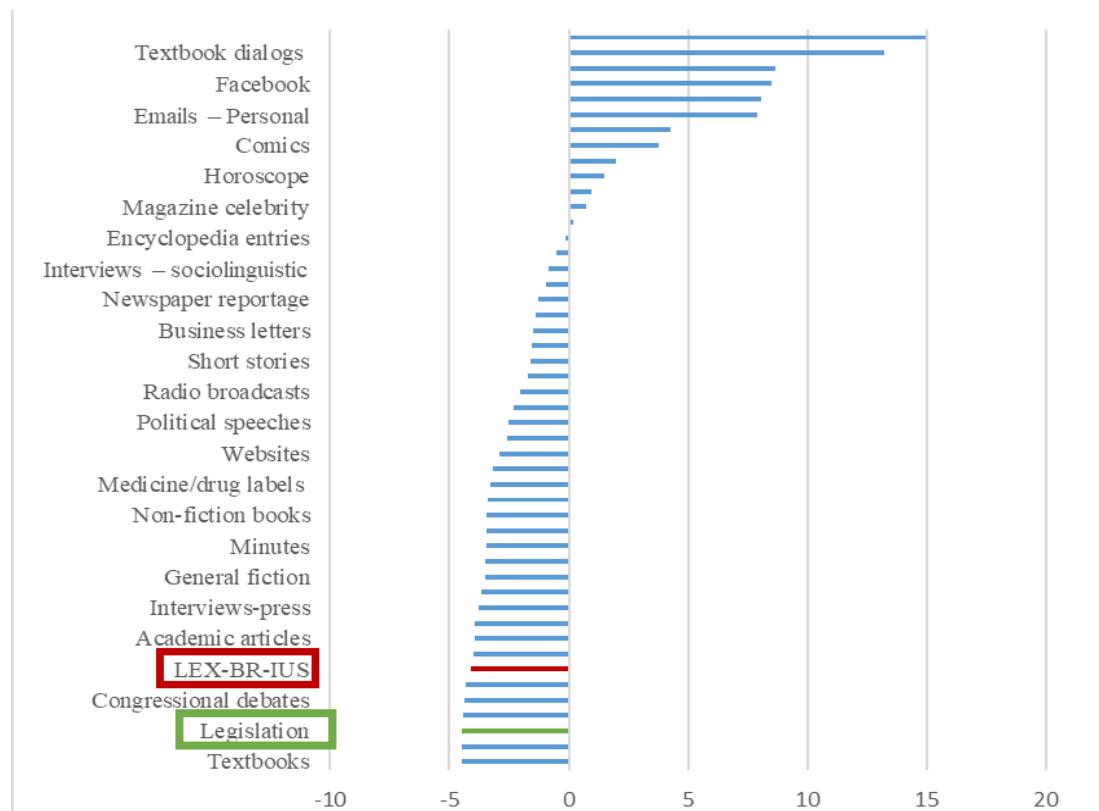


Figura 2. Dimensão 3

Como esperado, tanto nosso corpus quanto a seção “*Legislation*” pontuaram negativamente. Este polo apresenta apenas uma variável: relação *type-token*, que reflete a alta densidade lexical. Logo, seja nosso corpus que a seção “*Legislation*” apresentam pouca interação e alta carga informacional, tendo a última uma maior relação *type/token* que o primeiro. Como exemplo dessa característica nos nossos textos temos:

Art. 23. 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]

Neste exemplo podemos observar uma ampla gama de vocabulário especializado como: ilhotas, álveo, domínio público e domínio particular.

4.3 Dimensão 5: *Future versus past time orientation*

Na dimensão 5, por sua vez, temos, no polo positivo os registros cujo discurso é voltado para o futuro e, no polo negativo, aqueles voltados para o passado.

A seguir reproduzimos o gráfico dessa dimensão.

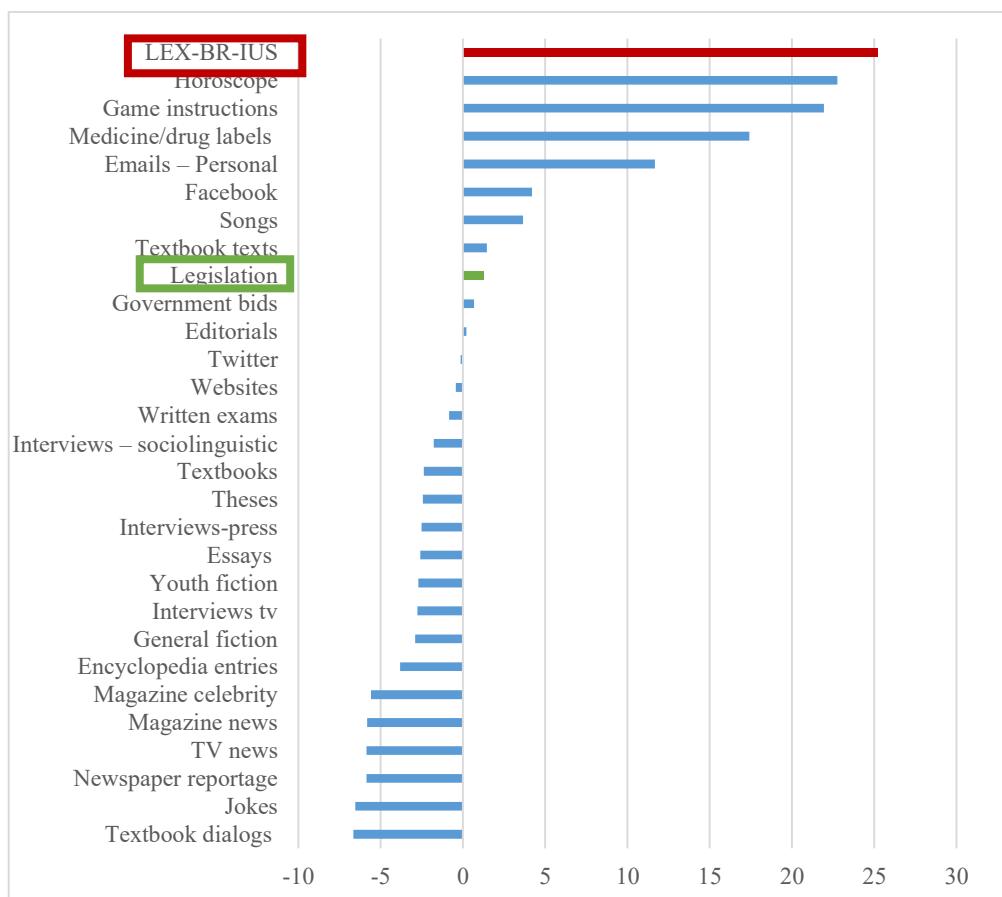


Figura 3. Dimensão 5

Essa é a dimensão na qual nosso corpus mais se destacou, pontuando superiormente a todos os registros do CBVR no polo positivo. Observa-se ainda, uma grande diferença entre a sua pontuação e aquela da seção “*Legislation*”, que é drasticamente inferior à nossa. O polo positivo é marcado por: alta frequência de verbos no futuro e uso dos modais dever e poder, conjunções coordenadas e orações subordinadas. São exemplos desses traços no nosso corpus:

Art. 8º § 2º O fornecedor deverá higienizar os equipamentos e utensílios utilizados no fornecimento de produtos ou serviços, ou colocados à disposição do consumidor, e informar, de maneira ostensiva e adequada, quando for o caso, sobre o risco de contaminação. [Brasil 1990]

No exemplo acima reproduzido observamos o uso do verbo modal dever no futuro e da conjunção coordenada “ou”, utilizados para prever e descrever os comportamentos exigidos dos destinatários da lei, assim como especificar e caracterizar os objetos desses comportamentos. De forma mais geral, o uso do futuro auxilia o legislador na descrição das normas, sua aplicação e consequências.

4.4 Significância estatística da variação

O resultado dos testes estatísticos empreendidos, quais sejam: ANOVA, p-value e R^2 indicam que nosso corpus se distingue significativamente dos registros do CVBR, inclusive “*Legislation*” em todas as dimensões analisadas. Já em relação à variação abarcada pelos textos que compõem o nosso corpus, também essa é estatisticamente relevante em todas as dimensões, capturando entre 53.4% a 81.6% da variação prevista para as dimensões do PB.

5 Discussão

O estudo de Berber Sardinha, Kauffmann e Acunzo (2014) é o mais completo sobre a variação do português brasileiro já realizado até o momento. O corpus utilizado –CVBR– foi compilado sob medida para a sua realização visando abranger o maior número possível de registros em PB de forma a obter uma descrição fidedigna da variação linguística nessa língua. Para tanto, o corpus amostra 48 registros orais e escritos, com 20 textos integrais por registro, totalizando 960 textos e aproximadamente 5.6 milhões de palavras. Dentre os registros que o compõe destacamos o “*Legislation*” composto por 20 textos legais que juntos totalizam 125 mil palavras (aprox. 2.2% do número de palavras do corpus). Tal seção é constituída por textos dos tipos normativos: Constituição, Códigos, Leis Ordinárias, Decretos, Medidas Provisórias e Atos Institucionais, sendo que alguns deles não são previstos pela atual Constituição (Atos Institucionais) e outros não estão mais em vigência. Não há também nenhuma separação interna na seção visando agrupar as diferentes espécies normativas. Cabe ressaltar que a maioria deles data de 2001, o que indicaria um recorte temporal, entretanto esta informação não consta da descrição do corpus. Da mesma forma, não foram disponibilizadas pelos autores informações importantes como os critérios adotados para a seleção dos textos, de onde foram extraídos ou quando, limitando-se os textos a terem dimensões similares àquelas das outras seções do CVBR.

A diferença, em alguns casos significativa, nos resultados obtidos entre a seção *Legislation* do CVBR e o *LEX-BR-Ius* parece-nos ser imputável ao nível de curadoria na seleção e limpeza dos textos que a compõem. O *LEX-BR-Ius* foi compilado respeitando o critério de atualidade e frequência de uso das normas legais [Ferrari e Marques 2022; Marques 2023], como apontamos acima, e com uma grande atenção na limpeza dos dados, conferida manualmente. Acreditamos que o registro “*Legislation*” não seja completamente representativo de todas as espécies normativas previstas na Constituição e inclui algumas não previstas, abarcando textos que não são mais aplicáveis e não tendo um critério de seleção claro. Além disso, da realização da limpeza não foram excluídas

as informações extratextuais como: índices, informações editoriais e números de página, o que pode enviesar os dados.

6 Conclusão

A AMD aditiva empreendida e seus resultados, juntamente com aqueles dos testes estatísticos, indicam que nosso corpus tem cargas fatoriais únicas e variação estatisticamente significativa em todas as dimensões analisadas. Tais dados nos levam a concluir que os textos legais são um registro segundo a perspectiva teórica por nós adotada, o que confirma nossa hipótese inicial. Com nosso trabalho esperamos ter contribuído na caracterização de um registro específico, aquele dos textos legais brasileiros. Estamos no momento completando o processamento dos dados do *LEX-BR-Ius* de maneira a poder realizar uma AMD completa no corpus e obter dados ainda mais refinados para uma descrição da linguagem utilizada nas normas legais do PB.

Agradecimentos

As autoras agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa (nº 88887.626989/2021-00) que permitiu esta investigação, à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelas bolsas de Iniciação Científica da equipe de compilação do corpus, e a Carlos Kauffmann pelo fornecimento dos dados do CVBR e etiquetagem do corpus.

Referências

- Barbera, M.; Onesti, C. (2009). Scheda Progetto di ricerca n. 9. *Corpus Jus Jurium*. In: Diadori, P (Ed.). *Progetto JURA: la formazione dei docenti di lingua e traduzione in ambito giuridico italo-tedesco*. Perugia: Guerra Edizioni, 2009. p. 349-351.
- Baker P., Hardie A. e McEnery T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh University Press.
- Berber Sardinha, T. (2010). A abordagem metodológica da Análise Multidimensional. *Gragoatá*, n. 29, p. 107-125.
- Berber Sardinha, T. (2013a). Variação entre registros da Internet. In: Shepherd, T. G.; Saliés, T. G. (Eds.). *Linguística da Internet*. Contexto, p. 55-85.
- Berber Sardinha, T. (2013b). *Pós-processador PT Tag Count*.
- Berber Sardinha, T., Kauffmann, C e Acunzo, C. M. (2014). Dimensions of register variation in Brazilian Portuguese. In: Pinto, M. V. (Eds.). *Multi-dimensional analysis: 25 years on a tribute to Douglas Biber*. John Benjamins Publishing Company.
- Biber, D. (1988). *Variations across speech and writing*. CUP.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, v. 8, n. 4, Oxford: Oxford University Press, p. 243-257.
- Biber, D e Conrad, S. (2009). *Register, genre, and style*. CUP.
- Bick, E. (2014). PALAVRAS, a constraint grammar-based parsing system for Portuguese. In: T. Berber Sardinha, T.; Ferreira, T. São Bento (eds.), *Working with Portuguese corpora*, p. 279–302. Bloomsbury.
- Bick, E. (2000). *The Parsing System "Palavras"*: Automatic Grammatical Analysis of

- Portuguese in a Constraint Grammar Framework. Tese (PhD) – Aarhus University. Aarhus, Denmark: Aarhus University Press.
- Brasil (1990). *Lei nº 8.078, de 11 de setembro de 1990* (Código de Defesa do Consumidor).
- Brasil (1934). *Decreto nº 24.643 de 10 de julho de 1934* (Código de águas).
- Brasil (2010). *Lei nº 12.288, de 20 de julho de 2010* (Estatuto da Igualdade Racial).
- Carapinha, C (2018). A linguagem jurídica. Contributos para uma caracterização dos Códigos Legais. *REDIS: Revista de Estudos do Discurso*, n. 7. Disponível em: <https://ojs.letras.up.pt/index.php/re/article/view/6200>. Acesso em: 10 set. 2021.
- Coulthard, M.; Johnson, A.(eds.) (2010). *The Routledge handbook of forensic*. Routledge.
- Coulthard, M.; Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Routledge.
- Ferrari, L. A.; Marques, C. G. de F. (2022) O *LEX-BR-Ius*: arquitetura e decisões na compilação de um corpus representativo das leis federais brasileiras. *ANTARES*, v.14, n.34. Disponível em: <http://www.ucs.br/etc/revistas/index.php/antares/article/view/11150/5328>. Acesso em: 19 dez. 2022.
- Ferrari, L. A.; Marques, C. G. de F. (em preparação) O *LEX-BR-Ius*: apresentação do corpus e pesquisas preliminares.
- Goźdź-Roszkowski, S. (2012). Legal Language. In: Chapelle, Carol A. (Org.). *The Encyclopedia of Applied Linguistics*. John Wiley e Sons, p. 3281-3287.
- Lorz R. A. (2019). Creating Law with Language – Crossing Borders and Connecting Disciplines from the Perspective of Legislative Practice. In: VogeL F. (ed). *Legal Linguistics Beyond Borders: Languag and Law in a World of Media, Globalisation and Social Conflict Relaunching the International Language and Law Association (ILLA)*. Duncker e Humblot GmbH, p. 5-8.
- Marques, C. G. de F. (2023). *Análise multidimensional dos textos legais federais brasileiros*. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte.
- McEnery, T. e Wilson A. (2001). *Corpus Linguistics*. Edinburgh UP, second edition.
- Tiersma, P. (1999). *Legal Language*. The University of Chicago Press.
- Tognini-Bonelli E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing Company.
- Sinclair, J. (2004). *Trust the Text: Language, Corpus, and Discourse*. Routledge.