

# ÍNDICES DE COMPLEXIDADE TEXTUAL EM SENTENÇAS DOS JUIZADOS ESPECIAIS CÍVEIS DO PODER JUDICIÁRIO DO ESTADO DO RIO GRANDE DO SUL

**Ester Motta**

<b>INFORMAÇÕES SOBRE A AUTORA</b>	
Ester Motta é graduada em Pedagogia pela Universidade Federal do Rio Grande do Sul (1986), onde realiza, atualmente, o doutorado em Letras. E-mail: <a href="mailto:estermottac@gmail.com">estermottac@gmail.com</a>	
<b>RESUMO</b>	<b>ABSTRACT</b>
Este artigo toma como pano de fundo as peculiaridades lexicais, gramaticais e estilísticas da linguagem jurídica, especialmente da linguagem empregada nas sentenças dos Juizados Especiais Cíveis do Poder Judiciário do Estado do Rio Grande do Sul. Tendo em conta que os Juizados Especiais foram criados para ampliar e facilitar o acesso do cidadão à Justiça, visa-se refletir sobre como, pela descrição linguística, poder-se-ia colaborar para promover a acessibilidade textual e terminológica dessas sentenças para um público leigo em Direito, o cidadão comum, com escolaridade limitada e pouco hábito de leitura. Para tanto, apresenta-se uma breve análise de índices de complexidade textual de um <i>corpus</i> de sentenças desses Juizados Especiais em comparação com textos de quatro jornais brasileiros: dois de referência - Zero Hora e Folha de São Paulo -, e dois populares - Diário Gaúcho e Jornal Massa. Por analisar textos de uma das áreas do conhecimento humano, o Direito, insere-se no âmbito dos estudos terminológicos, filiando-se, entre as vertentes teóricas da Terminologia como disciplina científica, à Terminologia Textual. Para a consecução das análises, valeu-se de algumas métricas da ferramenta Coh-Metrix-Dementia, criada por Cunha (2015), no âmbito da Linguística Computacional. Os resultados sugerem que as sentenças analisadas apresentam padrões lexicais que podem dificultar sua inteligibilidade para o seu público-alvo: pessoas de baixo poder aquisitivo, com pouca escolaridade e sem hábito de leitura.	The background of this article is the lexical, grammatical, and stylistic peculiarities of the legal language, especially the language used in judgments of the Special Civil Courts of the Judiciary of the State of Rio Grande do Sul. Considering that the Special Civil Courts were created to extend and facilitate the citizen's access to justice, it is intended to discuss on how, through the linguistic description, one could collaborate to promote the textual and terminological accessibility of these judgements to a lay audience, that is, the ordinary citizen, with limited education and little reading habit. Therefore, a brief analysis of textual complexity indices of a corpus of sentences of these Special Courts is presented and compared with texts of four Brazilian newspapers: two of them were considered as references - Zero Hora and Folha de São Paulo -, and two were considered popular newspaper - Diário Gaúcho and Jornal Massa. By analyzing texts from one of the areas of human knowledge, Law, it falls within the scope of terminological studies, being part of the Textual Terminology, along with the theoretical strands of Terminology as a scientific discipline. For the analyzes, some metrics of the tool Coh-Metrix-Dementia, created by Cunha (2015), in the scope of Computational Linguistics were used. Results suggest that the judgments analyzed have lexical patterns that can hinder its intelligibility for the target audience: people with low income, little education, and no reading habit.
<b>PALAVRAS-CHAVE</b>	<b>KEY-WORDS</b>
Acessibilidade Textual e Terminológica; Terminologia; Linguística Computacional.	Textual and terminological accessibility; Terminology; Computational Linguistics.

## INTRODUÇÃO

Neste estudo, fazemos uma análise contrastiva, num estudo qualitativo e quantitativo, entre textos de sentenças dos Juizados Especiais Cíveis (doravante JECs) do Poder Judiciário do Estado do Rio Grande do Sul (doravante PJRS) com textos de quatro jornais brasileiros: Zero Hora (doravante ZH), Folha de São Paulo (doravante FSP), Diário Gaúcho (doravante DG) e Jornal Massa (doravante JM). Considerando que os jornais escolhidos atingem várias camadas da população, o que propomos é um estudo comparativo a fim de verificar o quanto os padrões lexicais das sentenças dos JECs do PJRS são próximos ou distantes dos padrões lexicais dos textos jornais citados.

A preocupação com a questão de a linguagem jurídica ser mais acessível tem estado presente em várias esferas do Poder Judiciário brasileiro. É o caso, por exemplo, da campanha nacional pela simplificação da linguagem do Judiciário<sup>1</sup>, promovida pela Associação dos Magistrados Brasileiros (AMB), em 2005. Nessa campanha, a AMB editou um manual com explicações sobre o funcionamento da Justiça e com um glossário de expressões jurídicas. Na apresentação desse manual, assim se manifesta a AMB: “A Justiça deve ser compreendida em sua atuação por todos e especialmente por seus destinatários. Compreendida, torna-se ainda mais imprescindível à consolidação do Estado Democrático de Direito”. (AMB, 2007, p. 4)

No mesmo sentido, a Ministra Ellen Gracie Northfleet, em seu discurso de posse na Presidência do Supremo Tribunal Federal, no dia 27-04-2006, assim se manifestou:

Que a sentença seja compreensível a quem apresenta a demanda e se enderece às partes em litígio. A decisão deve ter caráter esclarecedor e didático. Destinatário do nosso trabalho é o cidadão jurisdicionado, não as academias jurídicas, as publicações especializadas ou as instâncias superiores. (NORTHFLEET, 2006, p. 1)

Assim, tendo em conta que a linguagem jurídica é tida como de difícil compreensão por grande parcela da população do país, eventuais desvios significativos dos padrões lexicais das sentenças analisadas poderão ser considerados fator que dificulte a compreensão por parte do cidadão que aos JECs do PJRS recorre.

Deve-se ter em conta também que, por representar uma ação linguística desenvolvida no âmbito do funcionamento jurídico, as sentenças dos JECs do PJRS apresentam uma linguagem referente a um determinado domínio do conhecimento humano, que, no campo dos estudos linguísticos, pode ser analisada pela ótica da Terminologia. Diante disso, faremos, a seguir, breves comentários acerca da Terminologia

<sup>1</sup>Disponível em <[http://www.amb.com.br/?secao=campanha\\_juridiques](http://www.amb.com.br/?secao=campanha_juridiques)>.

Textual, corrente teórica que embasa as análises do presente trabalho.

## 1 A TERMINOLOGIA TEXTUAL

No que se refere à Terminologia como disciplina científica, adotamos os princípios da Terminologia Textual, que, segundo Finatto (2004), nasce da compreensão de que o texto especializado está antes das análises lexicais, morfossintáticas, oracionais e frasais.

Esta autora cita os trabalhos de Lothar Hoffmann (2015), para quem há uma “precedência do objeto texto sobre o objeto termo”. Depreende-se daí “uma perspectiva de Terminologia textual que equivaleria a um *estudo do texto que tenha termos* e se distanciaria de um *estudo de termo sem textos*” (FINATTO, 2004, p. 352, grifo da autora).

Nos trabalhos de Hoffmann (2015), tem-se a noção de que “a especificidade das linguagens especializadas se expressa principalmente pela frequência de determinados recursos linguísticos, estatisticamente comprováveis” (FINATTO; KILIAN, 2015, p. 38). Essa noção remete aos trabalhos da Linguística de *Corpus* (doravante LC), que, segundo Zilio, apresentam pontos de convergência com a Terminologia Textual:

Em uma primeira abordagem, vemos que a Terminologia Textual entra com os métodos e aportes teóricos para o trabalho com textos especializados, enquanto a Linguística de *Corpus* entra com o aparato informático, com os pressupostos teóricos e com seu modo peculiar de observar a linguagem. (ZILIO, 2010, p. 138.)

Ciapuscio (1998), em artigo de que trata da terminologia do ponto de vista textual, analisou textos com a mesma temática dirigidos a leitores com perfis diferentes – especialista, semileigo e leigo –, e verificou que, nos textos destinados a semileigos e leigos, houve grande incidência de operações parafrásticas e variação formal da terminologia. Em trabalho posterior, Ciapuscio (2003) propõe uma categorização dos textos científicos com base no propósito do texto, nos seus interlocutores, que, por sua vez, determinarão os níveis de conteúdo semântico e formal-gramatical. Tais critérios se relacionam com a questão da complexidade textual, tema que embasa as análises de nosso estudo.

Diante disso, adotamos os pressupostos teóricos da Terminologia Textual, com os subsídios da LC, porque estudaremos a linguagem natural em uso sob a forma de um *corpus*, contando com apoio informatizado e tratamento estatístico. Adotamos igualmente os resultados dos estudos relativos à Complexidade Textual (doravante CT) e à Linguística Computacional (doravante LCOMP), que também entrarão com postulados teóricos e instrumentais para as análises que apresentamos neste artigo

Assim, considerando estabelecida a corrente terminológica à qual nos filiamos,

passaremos, a seguir, a abordar questões relativas à CT e à LCOMP.

## 2 A COMPLEXIDADE TEXTUAL E A LINGUÍSTICA COMPUTACIONAL

A CT, segundo Pasqualini (2012), é normalmente analisada pela ótica da leitura e do ensino da leitura e apresenta os seguintes encaminhamentos:

- a) Centrar-se em características ou elementos presentes em determinados tipos de textos e associá-los a determinadas dificuldades ou facilidades de compreensão de uma determinada categoria de leitores;
- b) Inferir o *modus operandi* do processamento mental cognitivo de diferentes tipos de leitor mediante aplicações de testes de compreensão após ou durante diferentes tipos de experimentos de leitura com diferentes tipos de textos.
- c) Reunir os indicativos dos itens a e b em prol da melhoria do ensino de leitura ou da produção de materiais informativos com maior acessibilidade de compreensão por leitores com proficiência de leitura reduzida ou em fase de desenvolvimento. (PASQUALINI, 2012, p. 46.)

Pasqualini (2012) cita vários autores, como Davison e Green (1988), segundo os quais, no âmbito internacional, desde 1920, há registros de pesquisas sobre *readability* ou inteligibilidade ou complexidade linguística, que surgiram pela necessidade de adequar materiais de leitura a públicos específicos.

Em relação ao Brasil, Perini (1982) foi um dos primeiros linguistas a estudar o tema da leitura funcional, da maior ou menor habilidade de leitura. Para este autor, os alunos deveriam ser expostos a textos compatíveis com seu nível de escolaridade e de compreensão.

Entre outros autores que também se dedicam a estudar este tema, encontra-se Leffa (1996), para quem o processo de compreensão leitora deve levar em conta três aspectos essenciais: o leitor, o texto, e as circunstâncias em que se dá o encontro entre ambos. Nesse sentido, cumpre referir que, em relação ao leitor, no Brasil, o Indicador de Alfabetismo Funcional (INAF), na edição de 2016<sup>2</sup>, aponta que apenas 8% dos brasileiros entre 15 e 64 anos conseguem elaborar textos mais complexos, interpretar tabelas e gráficos envolvendo mais de duas variáveis e resolver situações-problema de contextos diversos.

Deve-se ainda, no que diz respeito à CT, considerar as contribuições da LCOMP. Segundo Finatto et al. (2015), como a LC se vale de ferramentas e métodos computacionais para descrever, de modo estatístico, extensas amostras de usos da linguagem natural, um

---

<sup>2</sup>Disponível em: <<http://ipm.org.br/relatorios>>. Acesso em 15 fev. 2017. Pesquisa realizada em 2015 com 2 mil pessoas pelo Instituto Paulo Montenegro (IPM) e pela ONG Ação Educativa.



trabalho em parceria entre os cientistas da computação e os linguistas é processo que pode ser muito proveitoso. No que se refere ao trabalho dos linguistas, além de se valerem das informações extraídas pelas ferramentas da LCOMP, podem contribuir “para a melhoria desses sistemas, com a consideração de regras linguísticas que possam ser a eles integradas” (FINATTO et al., 2015, p. 47).

Assim, por acreditarmos nessa parceria entre os cientistas da computação e os linguistas, é que passaremos, na próxima seção, a tratar dos materiais e métodos para as análises empreendidas em nosso estudo.

### 3 MATERIAIS E MÉTODOS

Em nossa pesquisa, realizamos um estudo quantitativo e qualitativo sobre os padrões lexicais de um *corpus* de sentenças dos JECs do PJRS (*corpus* de estudo – CE) em contraste com um *corpus* de jornais tradicionais – ZH e FSP (Caso A) –, e com um *corpus* de jornais populares – DG e JM (Caso B).

As sentenças do CE representam as decisões tomadas em relação aos pedidos das pessoas que procuram os JECs para resolver algum conflito. São textos técnicos e, portanto, apresentam uma linguagem e configuração específica da área de que procedem. Para a constituição desse CE, coletamos, do *site* do TJRS<sup>3</sup>, quatro sentenças provenientes dos JECs e relativas a fevereiro de 2016. Após as prepararmos para análise pela ferramenta Coh-Metrix-Dementia, as salvamos num arquivo único denominado S1 – Sentenças 1.

Os jornais do Caso B – ZH e FSP – são considerados jornais de referência, por terem grande circulação no país, além de serem dirigidos a classes de maior poder aquisitivo, quais sejam, classes A, B e C. O primeiro, ZH, é publicado pela empresa jornalística RBS, em Porto Alegre, com versões impressa e *on-line*. O segundo, FSP, é editado na cidade de São Paulo, sendo o maior jornal de circulação do Brasil em formato digital e terceiro no formato impresso.

Os textos referentes à FSP foram extraídos do *site* <http://www.linguateca.pt/CETENFolha/e> fazem parte da Coleção CHAVE, criada pelo Linguateca em cooperação com a organização do CLEF – *Cross-Language Education and Function* – ([www.clef-campaign.org](http://www.clef-campaign.org)). Para o nosso estudo, em virtude do prazo de que dispúnhamos, só nos foi possível acessar os do ano de 1994.

Os textos da ZH foram extraídos de um bloco de 80 textos dos anos de 2006 e 2007, com um total de 55.528 palavras, utilizado pelas pesquisadoras Maria José B. Finatto, Carolina E. Scarton, Amanda Rocha e Sandra Aluísio em artigo em que procuraram

---

<sup>3</sup>Disponível em: <<http://www.tjrs.jus.br/site/>>. Acesso em 15-02-2017.

contrastar “a inteligibilidade de jornais populares e de jornais tradicionais brasileiros a partir de atributos coesivos, vocabulares e sintáticos, incluindo elipses” (FINATTO et al., 2011, p. 49). Este conjunto de textos pertence ao Projeto PorSimples, do NILC da USP de São Carlos e nos foi cedido para a pesquisa<sup>4</sup>.

Os jornais do Caso B – DG e JM – são populares, porque têm por objetivo oferecer notícias e informação de entretenimento e serviço às classes C, D e E. O primeiro, DG, é publicado pela empresa jornalística RBS, em Porto Alegre, com versões impressa e *on-line*. O segundo, JM, é publicado na Bahia, pelo grupo A TARDE, em versões impressa e *on-line*.

Os textos desses jornais foram extraídos do Site Porpopular<sup>5</sup>, que disponibiliza para *download corpus* desses dois jornais. O *site* Porpopular, coordenado pela Profa. Dra. Maria José Borcony Finatto, tem por foco a “descrição e o estudo de padrões do vocabulário exibido por textos de jornais populares voltados para públicos de menor poder aquisitivo” (FINATTO, 2017). Os textos do DG são referentes a meses variados do ano de 2013, e os relativos ao JM são referentes a maio de 2011.

Os textos dos jornais também foram arquivados em arquivos únicos, com a sigla do jornal a que se referem: ZH, para Zero Hora; FSP, para Folha de São Paulo; DG, para Diário Gaúcho; e JM, para Jornal Massa. Cabe-nos salientar que, apesar de os jornais pesquisados contemplarem regiões diferentes do Brasil – Sul, Sudeste e Nordeste –, e as sentenças serem provenientes dos JECs da Região Sul, neste estudo não foi analisada a questão dialetal.

A seguir, na Tabela 1, consta a quantidade de palavras dos *corpora* deste estudo.

Tabela 1 – Quantidade de palavras dos *corpora*

CORPUS DE ESTUDO	CASO A		CASO B		
	ZH	FSP	DG	JM	
Quantidade de palavras	8.294	8.616	8.125	8.335	8.313

Fonte: Produção da autora

Para a análise dos *corpora*, nos valem, dentre as ferramentas disponibilizadas pela LCOMP, do Coh-Matrix-Dementia, criada por Cunha (2015). Tal ferramenta nasceu no ambiente do Coh-Matrix-Port, que, por sua vez, é uma adaptação do sistema Coh-Matrix, que significa *cohesion metrics*. Trata-se de uma ferramenta

[e]laborada por pesquisadores da Universidade de Memphis, nos Estados Unidos (GRAESSER; McNAMARA; LOUWERSE; CAI, 2004), tem como propósito calcular índices de coesão e de coerência textual num amplo espectro de medidas lexicais,

<sup>4</sup> Gentilmente cedidos pela Profa. Dra. Maria José Bocorny Finatto.

<sup>5</sup> Disponível em: <[http://www.ufrgs.br/textecc/porlexbras/porpopular/download\\_do\\_corpus.php](http://www.ufrgs.br/textecc/porlexbras/porpopular/download_do_corpus.php)>. Acesso em 20-02-2017.

sintáticas, semânticas e referenciais com o fim de indicar a adequação de um texto a seu público-alvo (a “demanda cognitiva” e a legibilidade do texto). Também tem a função de apontar dados para identificar problemas textuais de ordem estrutural. (FINATTO, 2011, p. 39.)

No Brasil, o Núcleo Interinstitucional de LCOMP da USP – NILC – adaptou o Coh-Metrix em inglês para o português, o Coh-Metrix-Port. Esta adaptação ofereceu, inicialmente, das 60 métricas do sistema em inglês, apenas 35. O acesso é gratuito e se dá no âmbito do Projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital).

Tendo em vista que, no momento da pesquisa, o Coh-Metrix-Port não estava disponível para acesso, utilizamos para a nossa análise a ferramenta Coh-Metrix-Dementia, que disponibiliza as mesmas métricas que o Coh-Metrix-Port.

Na próxima seção, então, apresentaremos as análises empreendidas e os resultados delas advindos.

## 4 RESULTADOS

Em nosso estudo, utilizamos algumas das métricas selecionadas para as análises desenvolvidas na dissertação de mestrado de Mesquita (2011), pesquisa que tratou da legibilidade textual da Coleção Cadernos Didáticos de EJA, destinada a jovens e adultos egressos do ensino regular. Após a submissão de seus textos ao Coh-Metrix-Port, esta autora comparou os resultados obtidos com os do índice Flesch referentes aos mesmos textos.

No que se refere ao índice Flesch, ele é tido por muitos como uma fórmula superficial, pois considera apenas o número de palavras por sentença e o número de letras ou sílabas por palavra. Apesar disso, ele foi incorporado ao Coh-Metrix-Port e também ao Coh-Metrix-Dementia, pois, grosso modo, ele indica a complexidade na tarefa da leitura. Além disso, segundo Scarton e Aluísio (2010, p. 15), é a única métrica de inteligibilidade já adaptada para o português e incorpora o conceito de séries escolares brasileiras.

A fórmula para o cálculo do índice Flesch é:  $ILF = 164.835 - [1.015 \times (\text{Número de palavras por sentença}) - [84.6 \times (\text{Número de sílabas do texto}/\text{Número de palavras do texto})]$ . E o resultado é um número de 0 a 100, que, com base em Pasqualini (2012, p. 77), pode ser assim especificado: de 0 a 29 – muito difíceis; de 30 a 49 – difíceis; de 50 a 59 – razoavelmente difíceis; de 60 a 69 – textos padrão; de 70 a 79 – textos razoavelmente fáceis; de 80 a 89 – textos fáceis; e de 90 a 100 – textos muito fáceis.

As métricas selecionadas por Mesquita (2011) foram: *Índice Flesch; Número de palavras por sentenças; Número de verbos por número de sentenças; Sentenças por parágrafo; Sílabas por palavra de conteúdo; Incidência de negações; Frequências; Palavras antes dos verbos*

*principais; Incidência de conectivos; Sobreposição de argumentos adjacentes e Referência anafórica adjacente.*

Abaixo, colocamos a Tabela 2, com os dados de nossos *corpora* referentes a essas métricas.

Tabela 2 – Índices das Métricas do Coh-Metrix-Dementia Analisadas

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	S1	S2*	ZH-P	FSP	DG	JM
1 Sentenças por parágrafo x Palavras por sentença	44,824	42,787	28,954	33,672	30,867	38,546
2 Sílabas por palavra de conteúdo	2,943	2,955	2,872	2,829	2,785	2,733
3 Incidência de negação	10,731	11,244	7,984	11,216	12,019	11,535
4 Frequência de palavras de conteúdo	629308,3 86	449198,9 91	519824,3 56	436470,3 07	502725,0 80	523343,1 65
5 Cláusula por sentença	2,237	2,502	1,894	1,951	1,800	2,022
6 Incidência de conectivos	65,831	67,593	73,121	80,994	85,135	81,844
7 Palavras antes do verbo	4,005	4,767	4,755	4,575	4,618	5,144
8 Anáfora adjacente	0,184	0,231	0,124	0,240	0,151	0,238
9 Sobreposição de arg. adjacentes	1,056	1,186	0,395	0,398	0,169	0,546
10 Flesch	34,591	32,666	42,173	47,566	52,027	50,551

Fonte: Produção da autora

Na Tabela 2, acima, há uma coluna com S2 (Sentenças 2), não prevista na constituição dos nossos *corpora*. Fizemos isso porque, nas sentenças, usam-se julgamentos anteriores como argumentos para decidir, e esses julgamentos aparecem em forma de ementas jurisprudenciais. Essas ementas apresentam, inicialmente, palavras-chaves que resumem o assunto abordado. Assim, como nos textos das sentenças do CE aparecem muitas ementas, imaginamos que a presença de muitas frases nominais – às vezes, uma palavra só – poderia alterar os índices desses textos. Poderia alterar principalmente os relativos ao número de palavras por sentença, que, inclusive, é considerado para o cálculo do índice Flesch, para a aferição da inteligibilidade textual. Diante disso, fizemos um teste, criando esse outro arquivo – S2 –, sem o cabeçalho das ementas, para verificar se as diferenças entre o S1 seriam significativas. Tais situações procuramos descrever quando colocamos os resultados de nossas análises.

A seguir, então, fazemos uma aplicação das métricas selecionadas por Mesquita (2011) ao nosso estudo:

Quanto ao *Índice Flesch*, os textos do CE apresentaram-se mais difíceis, ficando os do S2 – 32,666 –, sem os cabeçalhos das ementas, com o índice menor que o S1 – 34,591. Isso veio a confirmar nossa suspeita de que a presença do cabeçalho das ementas poderia alterar a legibilidade desses textos.

Em relação ao *Número de Verbos por Sentença*, uma das métricas inseridas por Cunha



(2015) foi a *Cláusulas por Sentença*, que calcula o número médio de cláusulas por sentença, em que *cláusula* é caracterizada pela presença de um sintagma verbal. Quanto maior for o índice nesta métrica, maior é a incidência de verbos por sentença. Havendo maior incidência de verbos por sentença, pode-se inferir que haja um maior número de orações em cada sentença. Havendo maior número de orações, o período se torna mais complexo. No que diz respeito aos dados dos nossos *corpora*, os textos que se mostraram mais complexos quanto a esse aspecto foram os do CE – 2,237 (S1) e 2,502 (S2).

Mesquita (2011) afirma que *Número de Sentenças por Parágrafos* pode indicar se os parágrafos são longos ou curtos, o que pode estar relacionado ao grau de complexidade. Porém, cabe observar a orientação encontrada em muitos manuais de redação quanto a se preferirem frases mais curtas a longas. Assim, se um parágrafo é redigido com mais frases curtas, ele terá mais sentenças por parágrafo, mas nem por isso será mais complexo. cremos que, neste aspecto, o que mais importa é a relação entre o *número de palavras por sentença* e o *número de sentenças no parágrafo*, conforme consta na linha 1 da Tabela 2. Com base nesses cálculos, verificamos que os textos que apresentam os parágrafos com maior número de palavras e, em princípio, mais complexos, seriam os do CE – 44,824 e 42,787 –, seguidos dos textos do JM – 38,546.

Relativamente à métrica *Sílabas por palavra de conteúdo*, Mesquita (2011) refere que analfabetos funcionais apresentam dificuldade na decodificação do sinal gráfico. Segundo estudos relativos a esse tema, como os de Kleiman (2001), palavras de conteúdo desempenham um papel fundamental na compreensão leitora, pois os leitores fixam 80% do movimento dos olhos nesse tipo de palavra, o que não ocorre com as palavras funcionais. Quanto a esse aspecto, os dados dos nossos *corpora* demonstram que os textos que se apresentaram novamente mais difíceis foram os do CE – 2,943 e 2,955.

Em relação à métrica *Palavras antes dos verbos principais*, Aluísio e Scarton (2010, p. 53) afirmam que sentenças com muitas palavras antes do verbo principal sobrecarregam a memória de trabalho dos leitores. Em nossos *corpora*, os textos que apresentaram maior índice foram os do JM – 5,144 –, seguidos dos textos da ZH – 4,755 –, ficando em terceiro lugar os textos das S2 – 4,767.

Referentemente à *Incidência de Negações*, Mesquita (2011) menciona que, quanto mais negações num texto, mais difícil pode ser a compreensão. Em nossos *corpora*, os jornais populares apresentaram-se mais complexos neste quesito: DG – 12, 019; JM – 11,535; S2 – 11,244.

No que pertine à *Incidência de Conectivos*, cuja presença facilita a compreensão, porque as relações semânticas entre as sentenças são indicadas por esses elementos, os textos de nossos *corpora* que se mostraram mais difíceis foram os do CE – 65,831 e 67,593 –, e os mais fáceis, com índices mais elevados, foram os do Caso B – DG (85,135) e JM

(81,844).

Em relação à *Sobreposição de Argumentos Adjacentes*, os dados dos nossos *corpora* revelam que os textos do CE se apresentaram menos difíceis, com maiores índices. Uma possível explicação para isso pode estar no fato de que o CE é composto por textos especializados do mundo jurídico. Nesse âmbito, há orientações para que, em prol da precisão terminológica, caso haja necessidade, se repitam as palavras. É o que se vê no seguinte trecho de Ruy Rosado de Aguiar Júnior, Desembargador do TJRS e Ministro do Superior Tribunal de Justiça aposentado:

Se for preciso repetir as palavras, que estas sejam repetidas, isto é, não é preciso invocar sinônimos para expressar as idéias que estão na ementa. Especialmente, não se deve inventar: denúncia é denúncia, e não peça-ovo nem peça-pórtico; petição inicial é petição inicial, e não portal. Se tivermos de dizer dez vezes petição inicial, diremos sempre petição inicial, pois assim diz a lei, e é assim que devemos escrever. (AGUIAR JÚNIOR, 2008, p. 5.)

Quanto à *Referência Anafórica Adjacente*, se mais alto for o seu valor, mais difícil será o texto, porque haverá, em princípio, mais candidatos a serem retomados pela anáfora. Em nossos *corpora*, os textos com índice maior relativamente a esta métrica—mais difíceis, portanto— foram os dos Jornais FSP – 0,24 - e JM – 0,238.

A métrica *Frequência de Palavras de Conteúdo* indica a média de todas as frequências de palavras de conteúdo. Nesse aspecto, quanto maior a incidência de palavras não muito frequentes em um texto, maior a sua complexidade. Na pesquisa de Aluísio e Scarton (2010), elas observaram que os textos reescritos, mais fáceis, apresentaram um índice maior de frequência de palavras de conteúdo que os textos originais. Em nossos *corpora*, os textos que se apresentaram menos complexos neste quesito foram os das S1 – 629308,386. Aqui também parece calhar a observação feita quanto à repetição de palavras na métrica relativa à sobreposição de argumentos adjacentes.

A seguir, com base nas análises a que procedemos relativamente aos dados da Tabela 2, apresentamos a Tabela 3. Nessa tabela, consta a hierarquia de legibilidade de cada um de nossos *corpora*, tal como fez Mesquita (2011) em relação aos 13 cadernos da Coleção Cadernos Didáticos de EJA.

Tabela 3 – Hierarquia de legibilidade dos *corpora*

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	S1	S2*	ZH	FSP	DG	JM
1 Sentenças por parágrafo x Palavras por sentença	1	2	6	4	5	3
2 Sílabas por palavra de conteúdo	2	1	3	4	5	6

3	Incidência de negação	5	3	6	4	1	2
4	Frequência de palavras de conteúdo	1	5	3	6	4	2
5	Cláusula por sentença	2	1	4	5	6	3
6	Incidência de conectivos	6	5	4	3	1	2
7	Palavras antes do verbo	6	2	3	5	4	1
8	Anáfora adjacente	4	3	6	1	5	2
9	Sobreposição de arg. Adjacentes	5	6	2	3	1	4
	<b>Média dos itens 1 a 9</b>	3,555	3,111	4,111	3,888	3,555	2,777
	<b>Ordenação dos itens 1 a 9</b>	3	2	5	4	3	1
10	<i>Flesch</i>	34,591	32,666	42,173	47,566	52,027	50,550
	<b>Ordenação do Flesch</b>	2	1	3	4	6	5

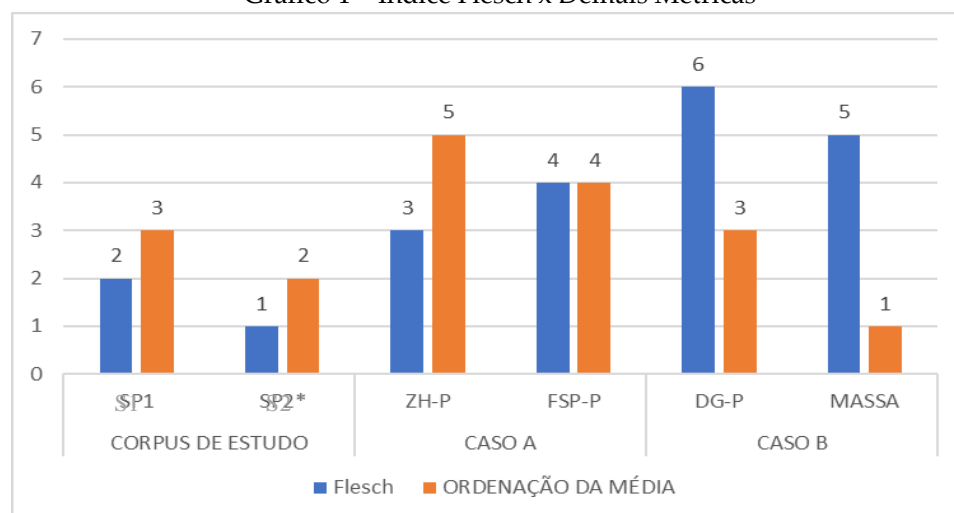
Fonte: Produção da autora

Para compreensão da Tabela 3, acima, a numeração de 1 a 6 corresponde a uma ordem de complexidade, sendo 1 o número atribuído ao texto mais complexo e 6 ao mais simples. Após cada arquivo receber um número de acordo com o critério antes apresentado, fizemos a média desses números, colocada na linha da *Média dos Itens 1 a 9*. Após, na linha seguinte – *Ordenação dos Itens 1 a 9*–, consta a ordenação dos textos segundo a média da linha anterior, sempre considerando o número menor para o texto mais complexo.

Após, na linha *Flesch*, constam os índices relativos a essa métrica, e, na linha subsequente – *Ordenação do Flesch*–, a ordenação dos nossos textos de acordo com essa métrica, igualmente considerando o número menor para o texto mais complexo.

Com esses dados, foi possível chegar ao Gráfico 1, abaixo, que ilustra melhor as discrepâncias havidas nessa comparação.

Gráfico 1 – Índice Flesch x Demais Métricas



Fonte: Produção da autora

Segundo o Gráfico 1, acima, de acordo com o índice Flesch, os textos mais fáceis são os dos jornais populares do Caso B, mas há uma enorme discrepância em relação à média das outras métricas analisadas. Vemos que o JM, pela média das métricas analisadas, ficou como o texto mais difícil. Como já mencionamos, o Índice Flesch considera apenas o número de palavras por sentença e o número de letras ou sílabas por palavra, e por isso é tido como superficial. Pelo gráfico acima, é possível inferir que efetivamente assim seja.

Por sua vez, as métricas analisadas neste estudo-piloto são de ordem lexical-sentença por parágrafo, sílabas por palavra de conteúdo, palavras por sentença, frequência de palavras de conteúdo –, sintática –incidência de negação, incidência de conectivos, número de palavras antes dos verbos – e semântica – anáforas, sobreposição de argumentos. Ou seja, as métricas analisadas contemplam um espectro maior de atributos do texto que o índice Flesch e podem, por isso, trazer uma imagem mais real acerca dos textos analisados. Porém, é preciso observar que as métricas *sentenças por parágrafo versus palavras por sentença e sílabas por palavra de conteúdo* conferem com os resultados trazidos pelo índice Flesch quanto a serem os textos do CE mais complexos que os demais.

Diante de todos esses dados, saltam aos olhos os índices dos textos jornalísticos, principalmente os dos jornais populares. Considerando que a atribuição de complexidade feita pelas métricas analisadas em nosso estudo, em comparação com os dados do índice Flesch, seja mais próxima da realidade dos textos de nossos *corpora*, causa-nos surpresa o fato de um jornal popular – JM – ficar como o mais complexo entre os textos analisados. Como já referimos, os jornais populares são destinados às classes C, D e E, com baixa escolaridade e pouco hábito de leitura. E, por terem esses leitores como destinatários de seus textos, em princípio, apresentariam uma linguagem mais simples, o que parece não estar acontecendo.

Nesse sentido, julgamos pertinente trazer as observações de Finatto et al. (2011), em artigo de que tratam de características do jornalismo popular. Segundo esses pesquisadores, os jornais populares brasileiros têm incrementado o índice de leitura de grandes parcelas da população urbana de baixa renda. Para tanto, afirmam, com base em Amaral (2006, p. 39), que uma pesquisa do IBOPE revelou que o DG foi responsável pela elevação do índice de leitores da região metropolitana de Porto Alegre para o maior de todo o Brasil.

Nesse artigo, para analisar os textos dos jornais, serviram-se das métricas do Coh-Matrix-Port e acrescentaram mais cinco métricas relativas a presença de elipses. E, pelos resultados obtidos, verificaram que o



DG, popular, exhibe, comparado ao tradicional ZH, similar uso de elipses. Portanto, se elipses equivalassem a texto mais complexo e até sofisticado em termos de elaboração, poder-se-ia detectar justamente aí um dos traços de um gênero novo, popular e ao mesmo tempo complexo”. (FINATTO et al., 2011, p.55-6)

Como vimos, os dados obtidos em nosso estudo parecem também favorecer a conclusão de estarmos diante de um “gênero novo, popular e ao mesmo tempo complexo”. (FINATTO et al., 2011, p.55-6)

## CONSIDERAÇÕES FINAIS

Cabe-nos, ao encerrarmos este estudo, salientar a importância dos recursos e ferramentas criados no âmbito da LCOMP que, mesmo não tendo por objetivo a pesquisa linguística, são de extrema valia. Dificilmente seria possível calcular todos os índices com os quais trabalhamos sem o aporte desse instrumental.

Por fim, cremos que as análises até aqui empreendidas permitem inferir que as sentenças dos Juizados Especiais Cíveis do Poder Judiciário do Rio Grande do Sul analisadas – sentenças dos CE – apresentam padrões lexicais que podem dificultar sua inteligibilidade para pessoas de escolaridade limitada e pouco hábito de leitura. Cumpre observar, porém, que os dados são relativos a uma pequena amostra de textos e a um número limitado de métricas, por isso os resultados daqui advindos não podem ser generalizados. Entretanto, revelam um potencial de análise textual bastante promissor para futuras pesquisas na área, como, por exemplo, estudos que promovam uma acessibilidade textual e terminológica em relação a essas sentenças. Com isso, estar-se-á contribuindo para a consecução de um dos objetivos dos JECs, qual seja, o de facilitar o acesso do cidadão ao Poder Judiciário.

## REFERÊNCIAS

AGUIAR JÚNIOR, Ruy Rosado de. Ementas e sua técnica. **Revista de Doutrina da 4ª Região**, Porto Alegre, n. 27, dez. 2008. Disponível em:

<[http://www.revistadoutrina.trf4.jus.br/artigos/edicao027/ruy\\_rosado.html](http://www.revistadoutrina.trf4.jus.br/artigos/edicao027/ruy_rosado.html)>. Acesso em: 13 abr. 2017.

AMARAL, Márcia Franz. **Jornalismo popular**. São Paulo: Contexto, 2006.

ASSOCIAÇÃO DOS MAGISTRADOS BRASILEIROS – AMB. O Judiciário ao Alcance de Todos – **Noções básicas de jurídiquês**. Disponível em:

<<http://www.amb.com.br/portal/juridiques/livro.pdf>>. Acesso em: 07 mar. 2010.

CIAPUSCIO, Guiomar Elena. **Textos especializados y terminología**. Barcelona: IULA, 2003.

CIAPUSCIO, Guiomar Elena. La terminología desde el punto de vista textual: selección, tratamiento y variación. **Organon**, Porto Alegre, v. 12, n. 26, p. 43-65, 1998.

CUNHA, André Luiz Verucci da. **Coh-Metrix-Dementia: análise automática de distúrbios de linguagem nas demências utilizando processamento de línguas naturais**. 2015. 155 f. Dissertação (Mestrado) – Curso de Ciências da Computação, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2015.

DAVISON, Alice; GREEN, Georgia M. (Ed.) **Linguistic complexity and text comprehension. Readability issues reconsidered**. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1988.

FINATTO, Maria José Bocorny. **Projeto PorPopular**. Disponível em: <[http://www.ufrgs.br/textecc/porlexbras/porpopular/download\\_do\\_corpus.php](http://www.ufrgs.br/textecc/porlexbras/porpopular/download_do_corpus.php)>. Acesso em: 20 fev. 2017.

FINATTO, Maria José Bocorny; LOPES, Lucelene; CIULIA, Alena. Processamento de Linguagem Natural, Linguística de *Corpus* e Estudos Linguísticos: uma parceria bem-sucedida. **Domínios da Linguagem**, Uberlândia, v. 9, n. 5, p.41-59, dez. 2015. Disponível em: <<http://www.seer.ufu.br/index.php/dominiosdelinguagem>>. Acesso em: 10 mar. 2017.

FINATTO, Maria José Bocorny e KILIAN, Cristiane Krause. Conceitos básicos da Lingüística de Linguagens Especializadas – texto comentário 2. In: Finatto, Maria José B.; ZILIO, Leonardo (Org.). **Textos e termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. Porto Alegre: Palotti, 2015. 256 p.

FINATTO, Maria José Bocorny; SCARTON, Carolina Evaristo; ROCHA, Amanda; ALUISIO, Sandra Maria. Características do jornalismo popular: avaliação da inteligibilidade e auxílio à descrição do gênero. In: VIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Cuiabá – MT. **Anais...** Cuiabá : Sociedade Brasileira de Computação, 2011. v. 1, p. 30-39.

FINATTO, Maria José Bocorny. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. **Organon** (UFRGS), v. 50, p. 30-45, 2011.

FINATTO, Maria José Bocorny. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, A.N.; KRIEGER, M. G. (Org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. Campo Grande: Ed. UFMS, 2004. v.2, p. 341-358.

KLEIMAN, A. B. (org.). **A formação do professor: perspectivas da linguística aplicada**. Campinas: Mercado de Letras, 2001.

HOFFMANN, Lothar. Conceitos básicos da Lingüística de Linguagens Especializadas. In: FINATTO, Maria José B. (Org.); ZILIO, Leonardo (Org.). **Textos e termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. Porto Alegre: Palotti,

2015. 256 p.

LEFFA, V. J. **Fatores da compreensão na leitura**. Projeto ELO, Ensino de Línguas Online: 1996. Disponível em: <[www.leffa.pro.br](http://www.leffa.pro.br)>. Acesso em: 20 fev. 2017.

MESQUITA, Nely Silva da Motta. **Legibilidade textual nos Cadernos Didáticos de EJA**: questões de processamento da leitura. 2011. 97 f. Dissertação (Mestrado). Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

NORTHFLEET, Ellen Gracie. **[Íntegra do discurso de posse da ministra Ellen Gracie como presidente do STF]**. Disponível em: <<http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=100112>>. Acesso em: 10 fev. 2016.

PASQUALINI, Bianca Franco. **Leitura, tradução e medidas de complexidade textual em contos da literatura para leitores com letramento básico**. 2012. 155f. Dissertação (Mestrado) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

PERINI, Mário Alberto. Tópicos discursivos e legibilidade. In: PERINI, M. A. (Coord.) **Definição lingüística da legibilidade**. Belo Horizonte: Relatório de Pesquisa ao Inep, 1982.

SCARTON, Carolina Evaristo; ALUÍSIO, Sandra Maria. Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. **Linguamática** (Revista para o Processamento Automático das Línguas Ibéricas), v. 2, n. 1, p. 45-61, 2010. Disponível em: <<http://linguamatica.com/index.php/linguamatica/article/viewfile/44/59>>. Acesso em: 20 fev. 2017.

ZILIO, Leonardo. Terminologia textual e lingüística de *corpus*: um estudo em parceria. In: PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO, Maria José Bocorny. **Linguagens especializadas em corpora**: modos de dizer e interface de pesquisa. Porto Alegre: Edipucrs, 2010. p. 184-201. Disponível em: <<http://www.pucrs.br/edipucrs>>. Acesso em: 15 fev. 2017.

## Título em inglês

# INDEXES OF TEXTUAL COMPLEXITY IN JUDGMENTS OF THE SPECIAL CIVIL COURTS OF THE JUDICIARY OF THE STATE OF RIO GRANDE DO SUL



# INVENTÁRIO