



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

## COMPARAÇÃO DE ÍNDICES DE COESÃO LEXICAL EM TEXTOS DO ENEM (2012-2022) NAS LÍNGUAS PORTUGUESA, ESPANHOLA E INGLESA

João Marcos Silvestre Sousa<sup>1</sup>; **Davi Alves Oliveira**<sup>2</sup>; Hernane Borges de Barros Pereira<sup>3</sup>

<sup>1</sup> Estudante do Curso de Licenciatura em Letras, Língua Inglesa e Literaturas (UNEB – DCH IV). IC Voluntário do PICIN/UNEB, joamarcossilvestre1@gmail.com;

<sup>2</sup> Doutorando no Programa de Pós-Graduação em Difusão do Conhecimento (PPGDC - UFBA, UNEB, IFBA, UEFS, LNCC, Senai Cimatec), Professor da Universidade do Estado da Bahia (UNEB), davioliveira@uneb.br.

<sup>3</sup> Doutor em Engenharia Multimídia, Centro Universitário Senai Cimatec & Universidade do Estado da Bahia, hbbpereira@gmail.com;

### RESUMO

Este artigo objetivou comparar índices de coesão lexical de textos em português, inglês e espanhol e verificar se tais índices variam conforme essas línguas. Analisamos textos das provas de Linguagens do Enem (2012-2022), considerados aqui como unidades de significado que obedecem a sete critérios de textualidade, destacando-se a coesão. Como método, o nosso estudo utilizou a análise de redes, modelando textos como redes semânticas de cliques e calculando índices de coesão lexical a partir delas. De acordo com os resultados, não há diferenças significativas entre os índices de coesão lexical das três línguas considerando os textos analisados, sugerindo, assim, que a coesão é um fenômeno linguístico geral que transcende as especificidades dessas línguas.

**PALAVRAS-CHAVE:** coesão lexical; coesão textual; redes semânticas.

<http://doi.org/10.55664/simbraredes2024.014>

### 1 INTRODUÇÃO

Um texto é uma unidade de significado<sup>1</sup> que obedece a sete critérios de textualidade<sup>2</sup> a saber, coesão (interconexão da superfície linguística em uma unidade), coerência (unidade lógica), intencionalidade (intenção do produtor do texto em manter a coerência), aceitabilidade (expectativa do receptor do texto em encontrar coerência), informatividade (nível de novidade apresentada ao longo do texto), situacionalidade (relação com o contexto) e intertextualidade (relação com outros textos). Destes critérios, a coesão é o que menos depende de características extratextuais e, por isso, é uma característica intimamente ligada à materialidade textual.

Diante desse cenário, o objeto de pesquisa deste estudo é a coesão textual, critério da textualidade que apresenta diversos recursos e se materializa por meio de diferentes elos coesivos, como a repetição lexical e o uso de pronomes e conjunções. Ademais, a coesão pode estar relacionada a uma coerência subjacente, auxiliando no desencadeamento lógico e organizado de ideias em um texto.

Para Hoey<sup>3</sup>, a coesão pode ser definida como a maneira em que certas palavras ou estruturas gramaticais conectam um período aos seus antecessores e sucessores em um texto (p. 3). Neste trabalho, consideramos como um período um conjunto de palavras delimitado pelos sinais de pontuação “.” (ponto), “...” (reticências), “!” (exclamação) ou “?” (interrogação) que possuem um significado completo. Além disso, nos concentramos em uma das classes de coesão estudadas por Hoey<sup>3</sup> e em Oliveira, Senna e Pereira<sup>4</sup>, a coesão lexical, uma vez que, segundo constatações do autor, ela é a classe mais recorrente em diversos textos, sendo um recurso recorrente na construção da coesão textual.

No presente estudo, partimos do pressuposto de que a coesão é uma característica da **superfície linguística**<sup>5</sup>, ou seja, ela é observada na materialidade textual e pode ser estudada sem necessariamente levar em consideração aspectos extralinguísticos. O estudo de Oliveira, Senna e Pereira<sup>4</sup> demonstrou que índices de coesão textual baseados em redes construídas a partir da superfície linguística conseguem diferenciar textos em língua inglesa de pseudotextos, isto é, conjuntos de períodos aleatórios. Considerando que as línguas diferem em sua superfície linguística, neste estudo investigamos a hipótese de que diferentes línguas também apresentam diferenças em seus padrões de coesão lexical. Se esta hipótese for corroborada por dados linguísticos, podemos considerar que a coesão é sensível a diferenças na materialidade textual percebida em diferentes línguas. Caso contrário, podemos manter a hipótese alternativa de que a coesão textual obedece a princípios gerais do funcionamento da língua que influenciam na materialidade textual de diferentes línguas.



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

Neste trabalho, a coesão lexical é estudada através de redes semânticas, estruturas formadas por elementos interconectados, chamados vértices, que podem ser conceitos, palavras ou lexemas, por exemplo (nos referimos a Pereira *et al.*<sup>6</sup> para uma revisão sobre definições de redes semânticas). Estas conexões são representadas por arestas que conectam os vértices. Utilizamos os índices de coesão lexical propostos por Oliveira, Senna e Pereira<sup>4</sup> para medir a coesão de textos escritos em português, espanhol e inglês encontrados em provas de Linguagens do Enem dos anos 2012 a 2022. O objetivo geral do estudo, então, foi verificar se há diferenças nos índices de coesão de textos das três línguas em questão. Como objetivo específico, o estudo se propôs a analisar a topologia de redes construídas a partir dos textos das três línguas para explorar outras possíveis características identificáveis nas redes relacionadas à coesão.

## 2 MATERIAIS E MÉTODO

Esta seção apresenta os materiais e o método utilizados no estudo, descrevendo os procedimentos de levantamento de dados, o método de construção das redes e as métricas consideradas nas análises.

### 2.1 Procedimentos de levantamento de dados

Com o objetivo de analisar os índices de coesão lexical em textos do Enem nas línguas portuguesa, espanhola e inglesa, primeiramente extraímos textos dos cadernos azuis das provas de Linguagens do Enem no período de 2012 a 2022. Em seguida, catalogamos informações dos textos, como ano da prova, língua do texto, número da questão em que o texto aparece, se atende aos critérios de seleção e, caso contrário, o motivo da exclusão. Estas informações estão disponibilizadas no repositório do projeto.<sup>7</sup> Em relação aos critérios de inclusão utilizados, o texto deveria ter as seguintes características:

- a) ser no mínimo verbal;
- b) apresentar mais de um período;
- c) ter os períodos marcados explicitamente por sinais de pontuação;
- d) não depender fortemente de imagens para completar seu significado;
- e) não depender fortemente de organização espacial para completar seu significado.

Foram elencados tais critérios uma vez que a coesão, enquanto possibilidades de conexão entre períodos por meio da superfície linguística, só poderia ser analisada em textos pelo menos verbais e com mais de um período. Seguindo essa mesma lógica, uma vez que as imagens e a organização espacial também contribuem para a construção semântica do texto, precisaram ser excluídos aqueles textos que dependessem fortemente delas, pois a análise de coesão lexical, utilizando o método baseado em redes, conforme proposto em Oliveira, Sena e Pereira<sup>4</sup>, não daria conta de tais textos.

Além disso, uma vez que nas provas de Linguagens do ENEM há mais questões com textos em português do que em espanhol e inglês, selecionamos os textos em português de forma em que a quantidade fosse equivalente à das outras línguas. Para exemplificar, na prova de 2018, foram selecionados 5 textos em inglês e 5 em espanhol. Logo, assim que os primeiros 5 textos em português foram escolhidos, não houve mais seleção naquela prova para que a quantidade de textos fosse proporcional entre as 3 línguas. No total, foram selecionados 96 textos, sendo 32 de cada língua.

### 2.2 Construção das redes

Para modelar textos com redes semânticas, primeiro separamos cada um em um arquivo de texto (.txt) e os limpamos manualmente, o que consistiu basicamente na remoção de delimitadores de períodos não utilizados como tal, por exemplo, reticências utilizadas para marcar uma elipse ao invés de delimitar um período. Depois, utilizamos o CohesionNet (versão 2.1.1)<sup>8</sup> para construir redes semânticas de cliques,<sup>9,10,11</sup> cujos vértices representam os radicais únicos das palavras do texto e as arestas representam a coocorrência dos respectivos radicais em um mesmo período, sendo cada período representado por uma clique (conjunto de vértices completamente conectados entre si). Foram consideradas radicais as palavras com suas flexões removidas a partir de um algoritmo de *stemming* utilizado pelo CohesionNet. Além disso, o programa calcula índices de coesão textual, implementando o método detalhado em Oliveira, Sena e Pereira.<sup>4</sup> Nas redes em questão, não são permitidas arestas múltiplas e as arestas não possuem peso nem direção.

Construímos uma rede para cada um dos textos para analisar os índices de coesão, totalizando 96 redes, 32 para cada língua, que chamaremos de “redes dos textos”. Comparamos os valores de coesão das três línguas quantitativamente a partir de testes ANOVA utilizando a linguagem de programação R<sup>12</sup>. Além das redes dos textos, construímos três redes a partir dos textos de cada língua para analisar a topologia de cada uma. Nos referimos a estas redes como “redes das línguas”. Por fim construímos uma rede aleatória



comparável às redes das línguas. A análise da topologia permite verificar possíveis diferenças e semelhanças na superfície linguística das línguas.<sup>13,14</sup>

### 2.3 Métricas utilizadas

Considerando cada rede de clique enquanto um grafo  $G = (V, E)$ , com  $n = |V|$  e  $m = |E|$ , analisamos, a partir das redes dos textos, os seguintes índices de coesão lexical conforme definidos em Oliveira, Senna e Pereira:<sup>4</sup>

- Índice Médio de Coesão Regressiva Global de Vértices ( $\langle\theta.\rangle$ ): média do nível de conexão de cada período com todos os outros períodos anteriores no texto;
- Índice Médio de Coesão Regressiva Local de Vértices ( $\langle\lambda.\rangle$ ): média do nível de conexão de cada período com o período imediatamente anterior no texto;
- Índice Médio de Coesão Pareada Média de Vértices ( $\langle\rho.\rangle$ ): média do nível de conexão de cada período com cada outro período presente no texto.

Além disso analisamos as seguintes métricas das redes das línguas:

- Grau médio ( $\langle k \rangle$ ): média dos graus dos vértices.
- Densidade ( $\Delta$ ): dada por  $\Delta = \frac{2m}{n(n-1)}$ ;
- Caminho mínimo médio ( $L$ ): média do tamanho dos caminhos mínimos existentes entre todos os pares de vértices conectados;<sup>15</sup>
- Coefficiente de aglomeração ( $C$ ): média da densidade dos subgrafos formados pelos vizinhos de cada vértice  $v \in V$ ;<sup>15</sup>
- Parâmetro de escala ( $\gamma$ ): parâmetro do ajuste da distribuição acumulada da probabilidade de grau dos vértices à lei de potência, calculado conforme descrito em Gillespie,<sup>16</sup> considerando que redes livres de escala apresentam distribuição de graus na forma  $P(K > k) \sim k^{-\gamma}$  e considerando  $p < 0,1$  para o teste de hipótese baseado em *bootstrapping* com 1000 simulações usando a estatística Kolmogorov–Smirnov.<sup>17</sup>

### 3 RESULTADOS E DISCUSSÃO

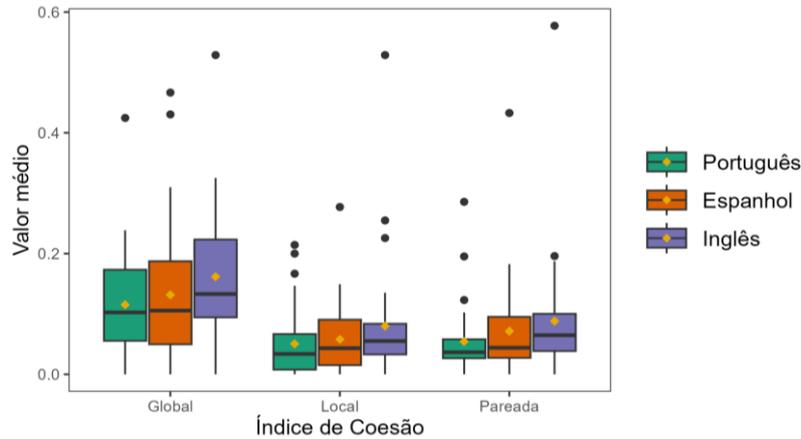
A Tabela 1 mostra a média ( $\mu$ ) e o desvio padrão ( $\sigma$ ) dos índices de coesão de cada língua. A Figura 1 ilustra os mesmos resultados, mostrando também a mediana da distribuição e presença de valores extremos. Como pode ser observado tanto na tabela quanto na figura, não há evidências de diferenças nos índices de coesão entre as línguas considerando os textos selecionados. A língua inglesa apresenta índices ligeiramente mais altos que as demais, mas esta diferença não é estatisticamente significativa (Global:  $F(2,93) = 1,6$ ;  $p = 0,207$ . Local:  $F(2,93) = 1,349$ ;  $p = 0,265$ ; Pareada:  $F(2,93) = 1,36$ ;  $p = 0,263$ ).

**Tabela 1.** Médias dos valores dos índices de coesão lexical calculados a partir das redes dos textos de cada língua

Língua	$\langle\theta.\rangle$		$\langle\lambda.\rangle$		$\langle\rho.\rangle$	
	$\mu^*$	$\sigma^*$	$\mu$	$\sigma$	$\mu$	$\sigma$
Português	0,115	0,088	0,050	0,057	0,054	0,058
Espanhol	0,132	0,118	0,058	0,058	0,071	0,081
Inglês	0,162	0,107	0,080	0,100	0,088	0,102

\*  $\mu$  = média;  $\sigma$  = desvio padrão.

**Figura 1.** Comparação entre os Índices de Coesão Global, Local e Pareada de textos em português, espanhol e inglês encontrados em provas do ENEM



Para a análise da topologia, inicialmente identificamos a quantidade de componentes em cada rede, resultando em 29 componentes na rede do português, 6 na do espanhol e 14 na do inglês. Por esse motivo, analisamos apenas o maior componente de cada rede, que é formado por 95% dos vértices na rede do português, 99% na do espanhol e 98% na do inglês. A Tabela 2 apresenta as métricas calculadas a partir dos maiores componentes das redes das línguas e da rede aleatória comparável e a Figura 2 ilustra as redes. A rede aleatória foi construída a partir do modelo  $G(n, p)$ <sup>18</sup>, sendo  $\bar{n} = 1230$  a média da quantidade de vértices das três redes das línguas e  $p$  a probabilidade de conexão entre os vértices, calculada considerando a média dos graus médios das três redes ( $\overline{\langle k \rangle} = 20,3$ ), igual a  $20,3, p = \frac{\overline{\langle k \rangle}}{\bar{n}-1}, 0,016$ . A rede aleatória resultante utilizada na análise contém apenas um componente.

Como ilustrado na Figura 2 e demonstrado na Tabela 2, as redes das línguas se assemelham em sua topologia ao apresentarem o fenômeno mundo-pequeno, ou seja, as três redes apresentam valores de  $L$  próximos e valores de  $C$  muito maiores aos da rede aleatória comparável. Por outro lado, a rede do inglês se difere das demais em relação a  $\gamma$ . Os ajustes das distribuições de grau à lei de potência das redes do português e do espanhol apresentaram valor de significância  $p < 0,1$ , o que significa que há evidência de que estas distribuições não seguem uma lei de potência (Figura 3). Por outro lado, a distribuição de grau da rede do inglês apresenta bom ajuste à lei de potência, mas apresenta valor de  $\gamma$  no intervalo  $3 < \gamma < 4$ , contrastando com os tipicamente observados, no intervalo  $2 < \gamma < 3$ .<sup>13</sup>

**Tabela 2.** Métricas calculadas a partir dos maiores componentes das redes das línguas e da rede aleatória comparável

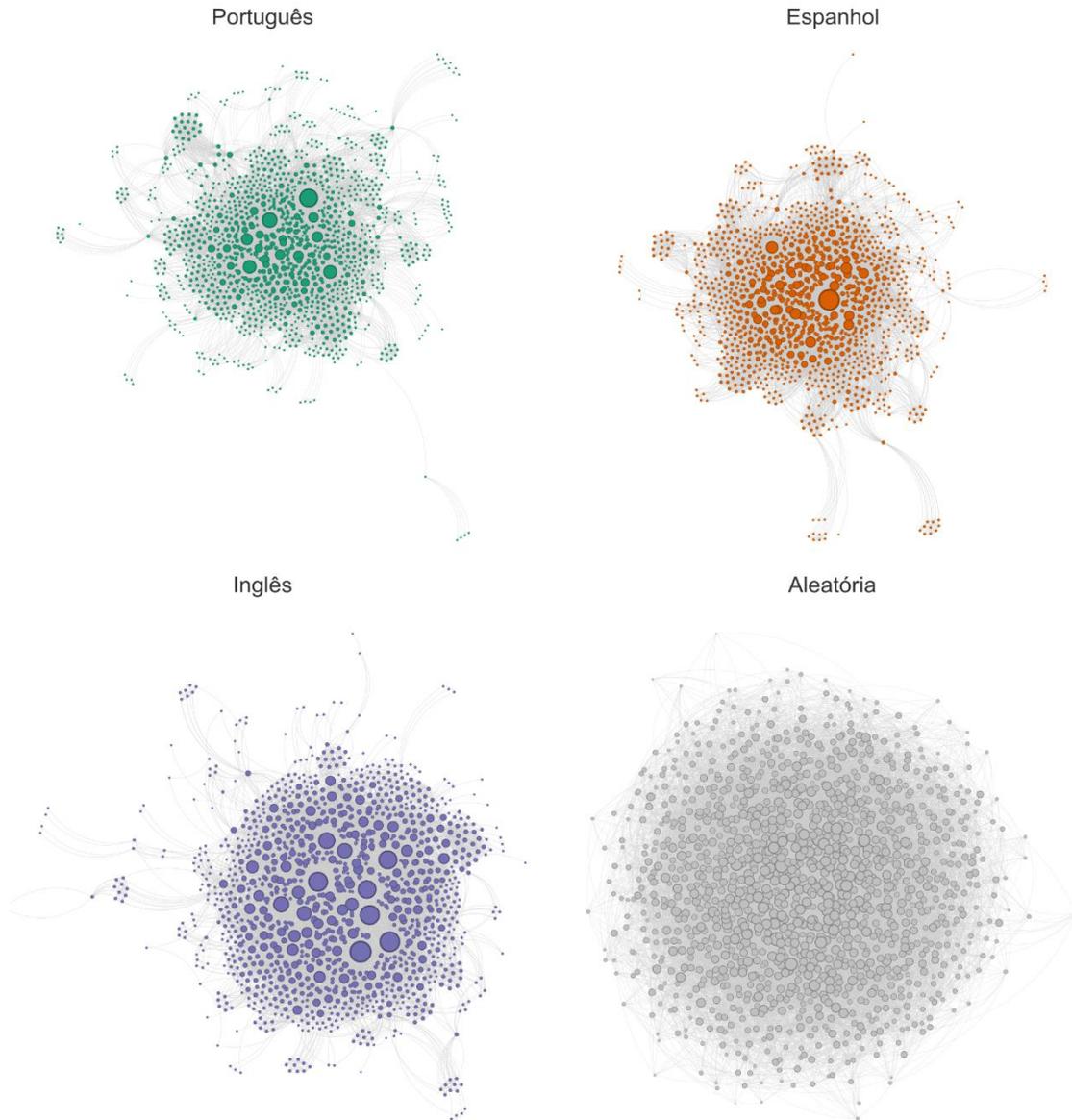
Rede	$n$	$m$	$\langle k \rangle$	$\Delta$	$L$	$C$	$\gamma$
Português	1264	11876	18,8	0,015	2,94	0,436	3.19*
Espanhol	1229	14586	23,7	0,019	2,75	0,436	2.97*
Inglês	1197	11064	18,5	0,015	2,96	0,415	3.94
Aleatória	1230	12473	20,3	0,016	2,69	0,017	--

\*  $p < 0,1$

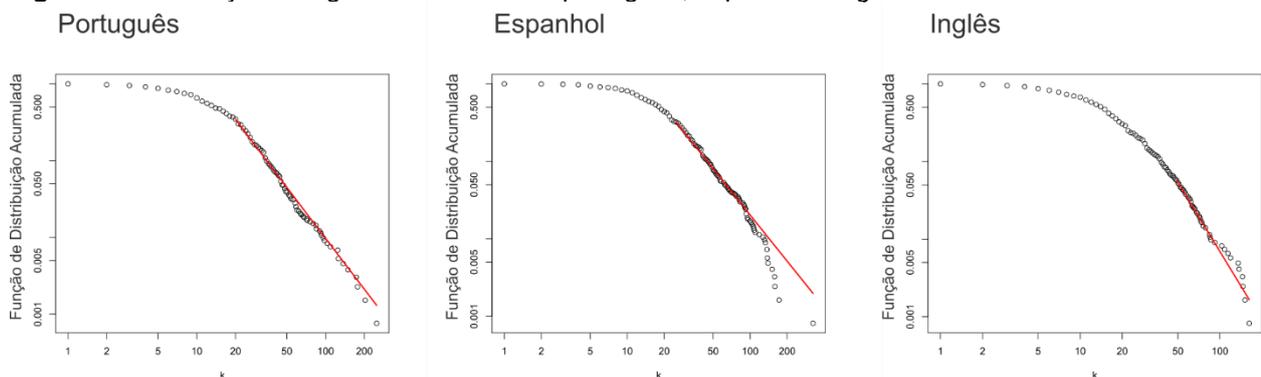


SimBraRedes  
III Simpósio Brasileiro de Teoria  
e Ciência de Redes

**Figura 2.** Maiores componentes das redes do português, espanhol e inglês, construídas a partir de textos do ENEM, e rede aleatória comparável



**Figura 3.** Distribuições de graus das redes do português, espanhol e inglês





SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

Os resultados evidenciam, em primeiro lugar, que a coesão, medida pelos índices de coesão lexical utilizados no presente estudo, é pouco afetada pelas diferenças entre as línguas portuguesa, espanhola e inglesa. Estudos com outras línguas e outros gêneros/tipos de texto são necessários para verificar se há padrões de coesão presentes nas línguas de modo geral, apontando para a coesão lexical enquanto uma característica linguística universal. Como o estudo de Oliveira, Senna e Pereira<sup>4</sup> demonstrou que esses índices de coesão são sensíveis aos gêneros textuais, uma interpretação possível para o resultado do presente estudo é de que a proximidade entre os índices se deu devido a alguma semelhança dos textos apresentados em provas do Enem, como o seu tamanho, mesmo em línguas diferentes. Esta interpretação leva à hipótese de que alguns gêneros textuais possuem índices de coesão característicos do gênero e que não são sensíveis às diferenças das línguas. Esta hipótese merece investigação em estudos futuros, pois aponta para características textuais perceptíveis a partir da materialidade linguística, mas que podem ser resultados de processos extralinguísticos.

Em segundo lugar, o resultado da análise da topologia das redes reforça a existência do efeito mundo pequeno em redes semânticas, conforme já demonstrado em estudos anteriores<sup>11,14,19,20,21,22,23,24</sup> (o estudo de Grabska-Gradzińska *et al.*<sup>25</sup> é uma exceção que merece atenção) e reforça resultados encontrados em Oliveira e Pereira<sup>11</sup> quanto à incerteza do comportamento livre de escala neste tipo de rede, contrastando com a maioria dos estudos mencionados.<sup>13,14,19,20,21,22,23,24</sup> Assim como em Oliveira e Pereira,<sup>11</sup> no presente estudo, o motivo do alto valor de  $\gamma$  para a rede do inglês não fica claro. Porém, duas observações lançam luz sobre este resultado: 1) Budel *et al.*<sup>13</sup> notaram desvios nas distribuições de graus observadas em comparação aos valores esperados por uma lei de potência em línguas altamente flexionais, que é o caso do português e do espanhol, e 2) Grabska-Gradzińska *et al.*<sup>25</sup> encontraram distribuições seguindo uma lei de potência com redes de textos literários, mas não em textos científicos. A interpretação mais plausível, considerando estes resultados, é que a predominância de flexões em português e espanhol resultam em distribuições de grau que desviam da lei de potência e textos do Enem. Além disso, apesar da diversidade de gêneros apresentados nas provas do ENEM, por serem selecionados para testar conhecimentos linguísticos, os textos, mesmo os de gêneros literários, podem possuir características típicas de textos científicos no que se refere à coesão e, por isso, também resultam em distribuições que desviam da lei de potência. Esta interpretação merece mais investigações, pois pode relevar que o comportamento livre de escala é sensível aos tipos ou gêneros textuais e outras características da superfície linguística, como predominância de flexões na língua.

#### 4 CONSIDERAÇÕES FINAIS

Constatamos que os índices de coesão dos textos nas 3 línguas (portuguesa, inglesa e espanhola) não apresentaram diferenças significativas no trabalho realizado. Uma razão para isso pode ser o fato de que as 3 línguas adotadas neste estudo estão próximas em relação às suas famílias linguísticas, sendo as três de raiz indo-europeia<sup>26</sup>, com o português e o espanhol de raiz itálica e com o inglês que, mesmo de origem germânica, foi fortemente influenciado pelo francês e pelo latim em sua evolução<sup>27</sup> e conseqüentemente compartilha muitos traços em comum com as outras duas línguas. Uma outra hipótese está relacionada à quantidade de textos, uma vez que, com mais textos, os resultados poderiam ser diferentes e, inclusive, variar conforme tipos e gêneros textuais. Uma terceira hipótese seria de que a coesão é uma característica universal da competência linguística humana, fazendo parte da Gramática Universal postulada por Chomsky<sup>28</sup>.

Diante disso, em pesquisas futuras, é possível testar tais hipóteses adicionando mais línguas a essa comparação, utilizando uma maior quantidade de textos, variando os tipos e gêneros textuais adotados e verificando as variações do comportamento livre de escala nas topologias de tais redes para uma análise mais aprofundada dos resultados contraditórios observados. De qualquer modo, o que os resultados deste artigo nos sugerem é que a coesão lexical é um fenômeno linguístico geral que transcende as especificidades das línguas analisadas.

#### Agradecimentos

Agradecemos ao Grupo de Pesquisa Formação em Linguagem e Ensino (FALE/UNEB) e ao Grupo de Pesquisa Fuxicos & Boatos (<https://redessociaiscomplexas.org/>) por possibilitarem momentos de formação que contribuíram para este trabalho e ao Programa Institucional de Iniciação Científica da UNEB (PICIN) pelo apoio durante a execução do projeto.

#### 5 REFERÊNCIAS

<sup>1</sup> HALLIDAY, M. A. K.; HASAN, R. **Cohesion in English**. London: Longman Group Limited, 1976.



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

- <sup>2</sup> BEAUGRANDE, R. A.; DRESSLER, W. **Introduction to Text Linguistics**. London: Longman, 1981.
- <sup>3</sup> HOEY, M. **Patterns of lexis in text**. Oxford: Oxford University Press, 1991.
- <sup>4</sup> OLIVEIRA, D. A.; SENNA, V.; PEREIRA, H. B. B. Indices of Textual Cohesion by Lexical Repetition Based on Semantic Networks of Cliques. **Expert Systems with Applications**, v. 237, n. 2024, p. 121580, 2024. <https://doi.org/10.1016/j.eswa.2023.121580>
- <sup>5</sup> TRAVAGLIA, L. C. A caracterização de categorias de texto: tipos, gêneros e espécies. **ALFA: Revista de Linguística**, v. 51, n. 1, p. 39–79, 2007. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/1>. Acesso em: 09 ago. 2024
- <sup>6</sup> PEREIRA, H. B. B.; GRILO, M.; FADIGAS, I. S.; SOUZA JUNIOR, C. T.; CUNHA, M. V.; BARRETO, R. S. F. D.; ANDRADE, J. C.; HENRIQUE, T. Systematic review of the “semantic network” definitions. **Expert Systems with Applications**, v. 1, n. 210, p. 118455, dez. 2022. <https://doi.org/10.1016/j.eswa.2022.118455>
- <sup>7</sup> OLIVEIRA, D. A. Comparação de índices de coesão lexical em textos do ENEM nas línguas portuguesa, espanhola e inglesa [Dados e Código] (1.2.0). Zenodo, 2024. <https://doi.org/10.5281/zenodo.13289106>
- <sup>8</sup> OLIVEIRA, D. A. CohesionNet (v2.1.1). Zenodo, 2024. <https://doi.org/10.5281/zenodo.11206287>
- <sup>9</sup> FADIGAS, I. S.; PEREIRA, H. B. B. A network approach based on cliques. **Physica A: Statistical Mechanics and its Applications**, v. 392, n. 10, p. 2576–2587, 2013. <https://doi.org/10.1016/j.physa.2013.01.055>
- <sup>10</sup> GRILO, M.; FADIGAS, I. S.; MIRANDA, J. G. V.; CUNHA, M. V.; MONTEIRO, R. L. S.; PEREIRA, H. B. B. Robustness in semantic networks based on cliques. **Physica A: Statistical Mechanics and its Applications**, v. 472, p. 94–102, 2017. <https://doi.org/10.1016/j.physa.2016.12.087>
- <sup>11</sup> OLIVEIRA, D. A.; PEREIRA, H. B. D. B. Modeling texts with networks: comparing five approaches to sentence representation. **The European Physical Journal B**, v. 97, n. 6, p. 77, jun. 2024. <https://doi.org/10.1140/epjb/s10051-024-00717-0>
- <sup>12</sup> R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2024. <https://www.R-project.org/>
- <sup>13</sup> BUDEL, G.; JIN, Y.; VAN MIEGHEM, P.; KITSACK, M. Topological properties and organizing principles of semantic networks. **Scientific Reports**, v. 13, n. 1, p. 11728, 20 jul. 2023. <https://doi.org/10.1038/s41598-023-37294-8>
- <sup>14</sup> FERRER I CANCHO, R.; SOLÉ, R. V.; KÖHLER, R. Patterns in syntactic dependency networks. **Physical Review E**, v. 69, n. 5, p. 051915, 26 maio 2004. <https://doi.org/10.1103/PhysRevE.69.051915>
- <sup>15</sup> WATTS, D. J.; STROGATZ, S. H. Collective dynamics of “small-world” networks. **Nature**, v. 393, n. 4, p. 440–442, 1998. <https://doi.org/10.1038/30918>
- <sup>16</sup> GILLESPIE, C. S. Fitting Heavy Tailed Distributions: The powerLaw Package. **Journal of Statistical Software**, v. 64, n. 2, 2015. <https://doi.org/10.18637/jss.v064.i02>
- <sup>17</sup> CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-Law Distributions in Empirical Data. **SIAM Review**, v. 51, n. 4, p. 661–703, 4 nov. 2009. <https://doi.org/10.1137/070710111>
- <sup>18</sup> ERDŐS, P.; RÉNYI, A. On random graphs. **Publicaciones Mathematicae**, v. 6, p. 290–297, 1959. <https://doi.org/10.5486/pmd.1959.6.3-4.12>
- <sup>19</sup> FERRER I CANCHO, R.; SOLÉ, R. V. The small world of human language. **Proceedings of the Royal Society of London. Series B: Biological Sciences**, v. 268, n. 1482, p. 2261–2265, nov. 2001. <https://doi.org/10.1098/rspb.2001.1800>
- <sup>20</sup> CALDEIRA, S. M. G.; PETIT LOBÃO, T. C.; ANDRADE, R. F. S.; NEME, A.; MIRANDA, J. G. V. The network of concepts in written texts. **European Physical Journal B**, v. 49, n. 4, p. 523–529, 2006. <https://doi.org/10.1140/epjb/e2006-00091-3>
- <sup>21</sup> FADIGAS, I. S.; HENRIQUE, T.; SENNA, V.; MORET, M. A.; PEREIRA, H. B. B. Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática. **Educ. Mat. Pesqui.**, v. 11, n. 1, p. 167–193, 2009. Disponível em: <https://revistas.pucsp.br/index.php/emp/article/view/2139>. Acesso em: 14 ago. 2024.
- <sup>22</sup> TEIXEIRA, G. M.; AGUIAR, M. S. F.; CARVALHO, C. F.; DANTAS, D. R.; CUNHA, M. do V.; MORAIS, J. H. M.; PEREIRA, H. B. B.; MIRANDA, J. G. V. Complex semantic networks. **International Journal of Modern Physics C**, v. 21, n. 3, p. 333–347, 2010. <https://doi.org/10.1142/S0129183110015142>
- <sup>23</sup> PEREIRA, H. B. B.; FADIGAS, I. S.; SENNA, V.; MORET, M. A. Semantic networks based on titles of scientific papers. **Physica A: Statistical Mechanics and its Applications**, v. 390, n. 6, p. 1192–1197, 2011. <https://doi.org/10.1016/j.physa.2010.12.001>
- <sup>24</sup> HENRIQUE, T.; FADIGAS, I. S.; ROSA, M. G.; PEREIRA, H. B. B. Mathematics education semantic networks. **Social Network Analysis and Mining**, v. 4, n. 200, p. 1–9, 30 dez. 2014. <https://doi.org/10.1007/s13278-014-0200-x>



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

- <sup>25</sup> GRABSKA-GRADZIŃSKA, I.; KULIG, A.; KWAPIEŃ, J.; DROŹDŹ, S. Complex network analysis of literary and scientific texts. **International Journal of Modern Physics C**, v. 23, n. 07, p. 1250051, jul. 2012. <https://doi.org/10.1142/S0129183112500519>
- <sup>26</sup> BALDI, P.; CUZZOLIN, P. Indo-European languages. *In*: COMRIE, B. (Ed.). **The world's major languages**. 3. ed. London and New York: Routledge, 2018.
- <sup>27</sup> BAKER, C. The effects of the Norman conquest on the English language. **Tenor of Our Times**, v. 5, artigo 5. Disponível em: <https://scholarworks.harding.edu/tenor/vol5/iss1/5>. Acesso em: 14 ago. 2024.
- <sup>28</sup> KENEDY, E. Gerativismo. *In*: MARTELOTTA, M. E. T. (Org.). **Manual de linguística**. São Paulo: Contexto, 2008, v. 1, p. 127-140.