TERMODINÂMICA DE OBRAS LITERÁRIAS

Meiryhelen Camila Domingues do Prado (PIBIC/CNPq/FA/UEPG), Prof. Dr. Sandro Ely de Souza Pinto, e-mail: desouzapinto@gmail.com.

Universidade Estadual de Ponta Grossa/Departamento de Física.

Física e Física Geral.

Palavras-chave: temperatura, processamento de linguagem natural, distribuição de Maxell-Boltzmann.

Resumo

Este trabalho tem como intuito apresentar um método para determinar a temperatura de obras literárias de Fernando Pessoa e seus heterônimos: Alberto Caeiro, Álvaro de Campos, Bernardo Soares e Ricardo Reis. Portanto, o texto é tratado como sistema físico, cujo suas palavras representam níveis de energia e a frequência de ocorrência destas, podem ser relacionadas a energia pela equação de Boltzmann.

Introdução

A Termodinâmica (do grego therme = calor e dynamis = movimento) é um ramo da Física, a qual estuda as variações de temperatura, pressão e volume de um sistema físico [1]. O conceito temperatura está diretamente relacionado com a energia cinética média das partículas, a agitação dessas partículas são associadas intuitivamente com o sentido de quente e frio [2].

Salienta-se a denotação “média” em energia cinética, em razão que temperatura é uma interpretação estatística dos movimentos microscópicos no interior dos corpos [3]. A área de estudo que é incumbida por explanar a conexão entre o mundo microscópico governado pelas leis da mecânica e a descrição macroscópica dos fenômenos físicos, é a mecânica estatística [4].

O cientista responsável pela relação entre termodinâmica e teoria cinética, foi o Físico austríaco, Ludwig Eduard Boltzmann (1844-1906). Com as evoluções que a Física teórica passou, em meados dos anos 1860 – 1870, pela introdução da Segunda Lei da Termodinâmica, Boltzmann prolongou a Teoria Cinética e desenvolveu a lei de equiparação de energia das partículas pelos graus de liberdade, e calculando quantas partículas possuem uma determinada energia, ocasionando a distribuição de Maxwell-Boltzmann. [5]

Nessa teoria, a probabilidade do sistema estar em um determinado microestado é inversamente proporcional à exponencial βE_n. Sendo $\beta = kT$, E_n a energia e n o nível do índice de energia; com k representando a constante de Boltzmann.

Baseando-se no trabalho Yamamoto e Miyazima [6] utilizamos a teoria de Maxwell-Boltzmann conforme o modelo descrito pelos autores.

Consideramos a probabilidade de ocorrência de uma lista de palavras de uma determinada linguagem, podendo associar tal probabilidade à energia de cada uma das palavras. Desta forma, calculamos a temperatura de um texto específico. Frisa-se que a lista de palavras provém da coleção de textos da língua considerada. Esta coleção denominamos corpus. O cálculo da temperatura é fornecido pela quantificação da diferença entre a distribuição de palavras do corpus e a distribuição de palavras do texto sob análise.

ISSN 1676-0093

Neste trabalho foi utilizado o Corpus Brasileiro [8] e as obras de Fernando Pessoa, e de seus heterônimos mais ilustres. Onde nossa análise não se estagnou somente em palavras, mas junto a estas, foi realizado separação de sílabas poéticas para melhores resultados.

A escolha desse autor em especifico, foi ocasionada por seus heterônimos distintos. Em distinção com Pseudônimo, que é um nome alternativo de um escritor, o heterônimo seria a criação de outra personalidades. Ou seja, ao realizar uma analise textual, os heterônimos possuem estilos de escritas diversos, aparentando ser outras pessoas.

Fernando Pessoa possuem em média 70 assinaturas de nomes diferentes em suas obras, todos com biografias e comportamentos individuais. Os mais conhecidos são Alberto Caeiro, Álvaro de Campos, Bernardo Soares e Ricardo Reis.

Material e Métodos

Primeiramente, foi realizada a busca do Corpus Brasileiro e das obras literárias disponíveis no domínio público [9] de Machado de Assis e de seus heterônimos: Alberto Caeiro, Álvaro Campos, Bernardo Soares e Ricardo Reis.

Com a coleta das obras, foi realizado a mineração de dados, onde foi retirado o desnecessário, como vírgulas e alguns pontos. A seguir, foi realizada duas separações para cada autor, onde uma delas seria em relação aos textos em prosa e a outra para poemas.

Para a listagem dos textos em prosa, foi necessário separar as palavras, onde cada uma uma foi disposta em uma linha. Já os poemas, foram escansionados com o auxílio do programa online Aoidos [10], onde cada sílaba poética se localizava em uma linha.

Com a listagem feita a cada autor, foi calculado a sua frequência com o programa R. A partir da frequência foi encontrado sua probabilidade, para por fim encontrar a energia de cada palavra a partir da relação de Maxwell-Bolztman. Dada a seguir:

$$P= \exp(-E/kT)$$
$$E=-\ln(P)$$

Para a obtenção da temperatura, utilizamos a energia do nosso corpus, Fernando Pessoa, e realizamos a correlação com os demais heterônimos.

A partir dos dados, foi realizado uma plotagem em gráfico. Em seguida, foi realizado a regressão para o ajuste da curva exponencial do gráfico, onde foi obtido o coeficiente de correlação e a temperatura que seguia a equação modelo:

$$y0+A1 \exp(-E/t1).$$

Onde modificamos a equação acima para:

$$y0= a0*\exp(-x/a1)$$

Onde para simplificações, a temperatura máxima que é dada em 100 K, foi considerada com t1= 1.

Foi realizado os mesmos procedimentos para os autores: Alberto Caeiro, Álvaro Campos e Ricardo Reis. Para a listagem em prosa e para a listagem de silabas poéticas.

Para a próxima analise, foi realizado os mesmos passos acima. Mas em contrapartida, foi realizado a utilização do corpus Brasileiro ao invés do Corpus Fernando Pessoa.

ISSN 1676-0093

Resultados e Discussão

As analises foram realizadas nos cinquenta primeiros elementos dos corpus, pois havia alguns elementos que não se encontravam no corpus fixado, e isto gerava pontos aleatórios nos gráficos.

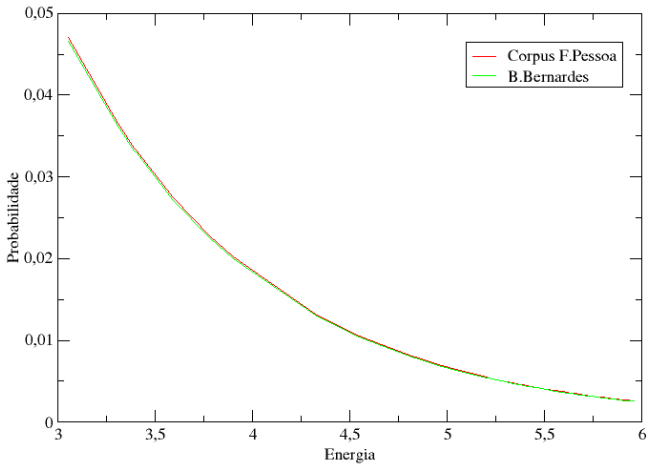
O modelo amostral da tabela utilizada para a obtenção da temperatura, segue abaixo com os dez primeiros elementos dos dados entre Bernardo Soares e a energia do corpus Fernando Pessoa (Tabela 1).

Tabela 1 – Lista de dados amostrais de Bernardo Soares em relação ao corpus Fernando Pessoa.

	Palav. B.Soares	Freq. Bernardo Soares	Energ. Fernando Pessoa
1	que	0,045698473922522	3,05646772764173
2	a	0,034986583934261	3,31018843265038
3	o	0,031244759349321	3,39475857521108
4	e	0,029389569008888	3,4013873873629
5	de	0,036873218178769	3,59584884376377
6	não	0,01859382860976	3,78175442659201
7	é	0,014610934093577	3,90118927744684
8	se	0,014893929230253	4,32602640862852
9	me	0,012713818547711	4,36261585606082
10	um	0,012504192520543	4,54916995994273

A partir dos dados retirados da tabela é plotado um gráfico entre a probabilidade do heterônimo Bernardo Soares e a energia do corpus Fernando Pessoa (figura 1).

Figura 01–Comparações das Probabilidades com função da Energia entre as distribuições do Corpus F.P. e de B. Soares. E o coeficiente de correlação.




```
        a0 = 0.99999
        a1 = 0.997154

Chi-square: 0.000318869
Correlation coefficient: 0.967397
RMS relative error: 0.579576
Theil U coefficent: 0.188678
```

Nossa temperatura é dada pelo elemento a1, obtido no gráfico.

Onde foi constatado que em relação ao corpus Fernando Pessoa, a temperatura dos heterônimos: Álvaro Campos, Alberto Caeiro, Bernardo Soares e Ricardo Reis, listados em palavras, variam em média 0,98 a 1,01 com um erro relativo de 0,57.

ISSN 1676-0093

Já os listados em escansões poéticas, a relação do corpus Fernando Pessoa com os seus heterônimos: Álvaro Campos, Alberto Caeiro e Ricardo Reis, variam em média 0,98 a 1,0 com um erro relativo de 0,7.

Os resultados obtidos em relação ao Corpus Brasileiro com os autores citados, variam em média 0,85 a 0,87 com um erro relativo de 0,5.

Conclusões

Com os dados obtidos podemos observar que Fernando Pessoa possui uma estilística com a mesma temperatura para seus heterônimos, mesmos que estes apresentem características temáticas e estilísticas diversas.

Em relação ao corpus Brasileiro, podemos obter a temperatura de suas obras literárias. Onde o autor Alberto Caeiro possui a maior temperatura de 0,877 em relação a este corpus.

Agradecimentos

Agradeço ao meu orientador profº Dr.º Sandro Ely de Souza Pinto, ao grupo de Dinâmica não Linear e Sistemas Complexos e a Fundação Araucária pelo auxilio financeiro.

Referências

[1] TERMODINÂMICA. Disponível em: https://mundoeducacao.bol.uol.com.br/fisica/termodinamica.htm. Acesso em: 22 ago. 2019.

[2] ALMEIDA, F. A. G; **Introdução a Física Estatística.** São Cristóvão-SE: CESAD, 2012. Disponível em: http://www.cesadufs.com.br/ORBI/public/uploadCatalago/09124110102012Introducao_a_Fisica_Estatistica_Aula_1.pdf. Acesso em: 22 ago. 2019.

[3]BARRETO,B.; XAVIER, C. **Física aula por aula.** Vol 2, 2.ed. São Paulo: FDT, 2013

[4]SILUO R. A. SALINAS . **HISTÓRIA DA MECÂNICA ESTATÍSTICA.** Cadernos de história e filosofia da ciência. São Paulo. Vol. 3. p. 28-42. 1982. Disponível em: https://www.cle.unicamp.br/eprints/index.php/cadernos/article/view/1170. Acesso em 23 ago. 2019.

[5]MILLAR, D. et al. **The Cambridge Dictionary of Scientists.** Cambridge University Press, 1996. Disponível em: http://www.fem.unicamp.br/~em313/paginas/person/boltz.htm. Acesso 23 ago. 2019.

[6] YAMAMOTO, K.; MIYAZIMA, S.: **MEASURING THE TENIPERATURE OF TEXT**, Fractals, vol 16, N01, pg 25-32 , 2008.

[7]SALINAS, S. **Introdução à Termodinâmica Estatística.** Instituto de Física da USP. 2016. Disponível em: **https://edisciplinas.usp.br/pluginfile.php/604072/mod_resource/content/1/notas%20de%20aula%20-%20vers%C3%A3o%202016-01.pdf.** Acesso 23 ago. 2019.

[8] CEPRIL: CORPUS BRASILEIRO. Disponível em <http://copusbrasileiro.pucsp.br/cb/Inicial.html> . Acesso em: 04 de junho de 2018.

[9]DOMÍNIO PÚBLICO. Disponível em <http://www.dominiopublico.gov.br/pesquisa/PesquisaObraForm.jsp> Acesso em: 11 de junho de 2018.

[10]AOIDOS. Disponível em: <https://aoidos.ufsc.br/>. Acesso em: 10 de junho de 2019.