UNIVERSIDADE FEDERAL DE VIÇOSA

MATHEUS AUGUSTO GONZAGA RODRIGUES

DESENVOLVIMENTO DE CORPORA E GERAÇÃO DE LETRAS DE MÚSICA UTILIZANDO MODELOS PRÉ-TREINADOS

VIÇOSA - MINAS GERAIS
2021

MATHEUS AUGUSTO GONZAGA RODRIGUES

DESENVOLVIMENTO DE CORPORA E GERAÇÃO DE LETRAS DE MÚSICA UTILIZANDO MODELOS PRÉ-TREINADOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Alcione de Paiva Oliveira

VIÇOSA - MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
R696d
2021

Rodrigues, Matheus Augusto Gonzaga, 1995-
Desenvolvimento de corpora e geração de letras de música utilizando modelos pré- treinados [recurso eletrônico] / Matheus Augusto Gonzaga Rodrigues. – Viçosa, MG, 2021.
1 dissertação eletrônica (42 f.): il. (algumas color.).

Orientador: Alcione de Paiva Oliveira.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 40-42.
DOI: https://doi.org/10.47328/ufvbbt.2021.015
Modo de acesso: World Wide Web.

1. Processamento de linguagem natural (Computação). 2. Transmissão textual. 3. Composição musical por computador. 4. Sistemas de coleta automática de dados. I. Universidade Federal de Viçosa. Departamento de Informática. Programa de Pós-graduação em Ciência da Computação. II. Título.

CDD 22. ed. 005.1

Bibliotecário(a) responsável: Alice Regina Pinto Pires CRB6 2523

MATHEUS AUGUSTO GONZAGA RODRIGUES

DESENVOLVIMENTO DE CORPORA E GERAÇÃO DE LETRAS DE MÚSICA UTILIZANDO MODELOS PRÉ-TREINADOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 06 de julho de 2021.

Assentimento:

Matheus Augusto Gonzaga Rodrigues
Autor

Alcione de Paiva Oliveira
Orientador

A Deus, minha família, aos orientadores
Alcione e Alexandra, e amigos.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me acolher em seus braços nos momentos que mais precisei, por guiar-me e apoiar-me diante de todas as situações adversas, não somente ao longo do processo do mestrado, mas também desde a graduação. Nada seria possível sem a fé Nele.

Aos meus pais, por me apoiarem em todas as decisões nesta fase de minha vida, por acreditarem no meu potencial e sempre olharem por mim diante das dificuldades. A minha mãe, principalmente, pelo fato de ter me ajudado e apoiado numa das minhas escolhas mais importantes da vida até então, que foi a de entrar no mestrado.

À minha namorada, que sempre me apoiou nos momentos difíceis, pela parceria de sempre, por cultivar sonhos e querer crescer juntamente comigo. Agradeço pelos conselhos, por aquela palavra amiga que precisei em incontáveis situações, por me motivar e por trazer o melhor de mim.

Aos meus amigos pelos conselhos, pela parceria nos momentos que eu me via sem motivação e sem forças para continuar.

Agradeço também à Universidade Federal de Viçosa (UFV), pela estrutura, pela qualidade das disciplinas e dos professores, em destaque ao meu orientador Alcione de Paiva e sua esposa e também professora da pós-graduação em Ciência da Computação, Alexandra Moreira, que muito me ajudaram no processo de obtenção deste título. Deixo aqui também meus agradecimentos aos demais professores do Departamento de Informática (DPI), em especial aos professores André Gustavo, Lucas Vegi, Vladimir Di Iorio e Levi Lelis, com os quais eu mais absorvi conteúdo, não só das disciplinas em si, mas também da profissão no geral.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

RESUMO

RODRIGUES, M., M.Sc., Universidade Federal de Viçosa, julho de 2021. **Desenvolvimento de Corpora e Geração de Letras de Música Utilizando Modelos Pré-treinados**. Orientador: Alcione de Paiva Oliveira.

O campo de geração de linguagem natural consiste na criação de textos que fornecem informações contidas em outros tipos de fontes (dados numéricos, gráficos, taxonomias e ontologias ou mesmo outros textos), com o objetivo de tornar esses textos indistinguíveis, na medida do possível, daqueles criados por humanos. A geração automática de texto possibilita o aumento da produção de material textual que pode ter diversas finalidades, tais como, produção de material didático, produção de manuais técnicos, auxílio na produção de material de divulgação científica, geração automática de propaganda etc. Dentro do escopo desta tarefa, destaca-se o gênero textual “letra de canção” que se caracteriza por sua estrutura (estruturada em versos que se agrupam em estrofes), por possuir rima e ritmo, e por ser capaz de despertar emoções, visto que o gênero pertence ao domínio artístico. Devido a essas características, a produção de texto musical apresenta desafios adicionais em relação à produção de textos em geral. A geração de letras de canções de forma automática pode auxiliar artistas em suas composições, reduzindo o tempo gasto na escrita de músicas e fomentando a produção musical. A proposta desta pesquisa é verificar a viabilidade da geração de texto musical por meio dos modelos mais recentes de aprendizado profundo. Para atingir esse objetivo a pesquisa foi realizada em duas etapas. A primeira consistiu no desenvolvimento de um *corpus* de letras de música para treinamento e/ou *fine tuning* de modelos de aprendizado. A segunda etapa consistiu no ajuste de um modelo pré-treinado para geração de letras de música. O resultado da pesquisa gerou evidências para a abordagem adotada, mostrando que é possível caminho promissor para este tipo de tarefa.

Palavras-chave: Processamento de linguagem natural. Geração de texto. Letras de música.

ABSTRACT

RODRIGUES, M., M.Sc., Universidade Federal de Viçosa, July, 2021. **Corpora Development and Lyrics Generation Using Pre-trained Models**. Adviser: Alcione de Paiva Oliveira.

The field of natural language generation consists in the creation of texts that provide information contained in other types of sources (numerical data, graphics, taxonomies and ontologies or even other texts), in order to make these texts indistinguishable, as far as possible, of those created by humans. The automatic generation of text makes it possible to increase the production of textual material that can have different purposes, such as the production of teaching material, production of technical manuals, assistance in the production of scientific dissemination material, automatic generation of advertisement, etc. Within the scope of this task, the textual genre "song lyrics" stands out, characterized by its structure (structured in verses that are grouped into stanzas), by having rhyme and rhythm, and by being able to arouse emotions, once that the genre belongs to the artistic domain. Due to these characteristics, musical text production presents additional challenges compared to text production in general. The automatic generation of song lyrics can help artists in their compositions, reducing the time spent writing songs and promoting music production. The purpose of this research is to verify the feasibility of generating musical texts through the most recent models of deep learning. To achieve this goal, the research was carried out in two stages. The first consisted in the development of a *corpus* of song lyrics for training and/or fine tuning of learning models. The second step consisted of adjusting a pre-trained model for song lyrics generation. The research results generated evidence for the adopted approach, showing that it is a possible promising path for this type of task.

Keywords: Natural language processing. Text generation. Lyrics.

LISTA DE FIGURAS

Artigo 1

Artigo 2

LISTA DE TABELAS

Artigo 1

Artigo 2

LISTA DE QUADROS

Artigo 2

LISTA DE SIGLAS E ABREVIATURAS

NLP	Natural language processing.
PLN	Processamento de Linguagem Natural.
GPT-2	Generated Pretrained Transformer 2.
URL	Uniform Resource Locator.
API	Application Programming Interface.
NLG	Natural Language Generation.
LSTM	Long Short-Term Memory.
GRU	Gated Recurrent Units.

SUMÁRIO

1. INTRODUÇÃO

A produção de textos é uma das formas de expressão mais utilizadas por nós, seres humanos. A mesma pode ser expressa em diversas línguas, em diversos formatos, seja ela de forma manual, digital, dentre outros meios. Mais ainda, os seres humanos se expressam através da escrita de maneiras totalmente distintas, visto que a mudança de contexto, do período de tempo no qual foi veiculado, do veículo ou meio de comunicação envolvidos, do estilo de escrita do escritor, seja por gênero textual ou grau de intelectualidade.

Armando Vieira (2011), doutor em Física Teórica pela Universidade de Coimbra, afirma que "a escrita é uma tarefa mal amada, senão mesmo menosprezada, por muitos cientistas e engenheiros que preferem aplicar seu tempo noutras atividades consideradas mais úteis." Embora, segundo o autor, a escrita seja uma tarefa "menosprezada", acredita-se que também é uma das principais formas de construção e divulgação de ideologias, comunicação, e, em diversas situações, se mostra como forma de expressão artística.

Carneiro da Silva (2013) acredita que o texto pode ser caracterizado de três formas diferentes, dado que, é um produto do pensamento, um simples instrumento de comunicação, mas também é um processo de interação entre autor e leitor. A autora ainda completa que, no primeiro caso, o autor se torna um sujeito psicológico, responsável por transmitir suas intenções no papel, fazendo com que o leitor consiga captá-las.

Nesse sentido, ganha-se destaque o gênero textual "letra de canção". Santos (2015) afirma que o gênero, também descrito como "letra de música", desenvolve a compreensão e a produção textual, além do fato de que, ao se ter contato com o gênero, o escritor e o leitor são capazes de despertar emoções e pensamentos críticos, visto que o gênero pertence ao domínio artístico. A autora ainda reforça a presença da estrutura musical, geralmente definida por estrofes e versos, métrica, rima, ritmo, dentre outras características linguísticas.

Acredita-se que o processo de escrita para um ser humano pode ser demorado, por vezes maçante, independente do gênero textual envolvido. De forma similar, a tarefa de transformar sentimentos, experiências, opiniões, ou qualquer outra abordagem em letras de música, também é uma tarefa que merece um(a) estudo/análise mais aprofundado(a). A proposta desta pesquisa é verificar a viabilidade da geração de texto musical por meio dos modelos recentes de aprendizado profundo. Para atingir esse objetivo a pesquisa foi realizada em duas etapas. A primeira consistiu no desenvolvimento de um *corpus* de letras de música para treinamento e/ou fine-tuning de modelos de aprendizado. A segunda etapa consistiu no ajuste de um modelo pré-treinado para geração de letras de música. O resultado da pesquisa gerou evidências para a abordagem adotada, mostrando que é possível caminho promissor para este tipo de tarefa.

1.1. Objetivos

A Linguística Computacional e a Linguística de *Corpus* são áreas que estão despertando grande interesse de pesquisadores e grandes corporações e as ferramentas de manipulação textual têm experimentado uma grande evolução.

Relacionada a essas áreas, a geração de textos em linguagem natural é uma tarefa de grande relevância. Pretende-se desenvolver, através da utilização de modelos pré-treinados, a geração de texto automática de cunho artístico, mais precisamente de letras de músicas, com apoio de uma base textual pré-coletada (autoral) de letras de músicas. Especificamente, pretende-se aplicar a tarefa de Automatic Text Generation no domínio artístico através de modelos linguístico-computacionais.

1.2. Objetivos específicos

Disponibilizar em algum site ou repositório de dados, tal como Common Crawl e/ou Kaggle um "*Corpus* musical" que se destaque em termos de volume se comparado com outros *Corpus* de mesmo propósito na literatura, e utilizável para a comunidade de linguística computacional (possivelmente para outras tarefas de PLN, além da geração automática de texto). É de objetivo específico também adaptar e avaliar a qualidade de um modelo pré-treinado para geração de texto de cunho artístico através de métricas de análise de coesão semântica.

1.3. Organização da dissertação

O trabalho apresentado aqui, foi elaborado através de uma coletânea de artigos desenvolvidos ao longo do processo de pesquisa.

Na primeira seção desta obra, apresenta-se uma análise da compilação de um *Corpus* gerado a partir de conteúdo da Web utilizando a técnica de Web Scraping, o que culminou no artigo "Development of a Song Lyric *Corpus* for the English Language" (2019), onde é explicado todo processo de extração do conjunto textual, seu tratamento e posteriores análises, dentre elas, N-grams analysis e Embedding. O artigo foi publicado no NLDB2019 (24th International Conference on Applications of Natural Language to Information Systems).

O trabalho citado acima trouxe a necessidade de verificar a possibilidade de gerar textos sintática e semanticamente corretos de forma automática em linguagens como português e em inglês, o que é discutido na segunda seção deste documento por meio do artigo "Geração de Letras de Música apoiada por Modelos Pré-treinados", no qual é utilizado um modelo de rede neural GPT-2 pré-treinado com apoio de Corpora gerada no artigo citado anteriormente. Mais ainda, foi abordada a validação da qualidade dos textos gerados utilizando técnicas como o cálculo da perplexidade, comparações entre a geração textual para diferentes linguagens dada à qualidade dos corpora utilizados para *fine-tuning* e uma análise mais profunda no que tange à exigência de recursos computacionais de alto desempenho para tal tarefa.

Por fim, são apresentadas as conclusões gerais do trabalho, acompanhadas de comentários referentes aos resultados alcançados e da abertura para discussões de possíveis trabalhos futuros.

2. ARTIGOS CIENTIFICOS

2.1. Development of a Song Lyric *Corpus* for the English Language

Rodrigues, Matheus & Oliveira, Alcione & Alexandra, Moreira. (2019). **Development of a Song Lyric *Corpus* for the English Language**. 10.1007/978-3-030-23281-8_33.

Abstract. *Web Scraping Tools are simplifying the task of creating large databases for various applications such as the construction of corpus aimed at the development of applications for natural language processing. Many of these applications require a large amount of data, and in that sense, the Web presents itself as an important data source. Among the various tasks in the NLP scope, one of the most challenging is automatic text generation. In this task the objective is to generate syntactically and semantically correct texts after a training process on a particular corpus. This article presents the elaboration of an English song lyrics Corpus, extracted from the Web, that can be used to train applications for automatic generation of lyrics, poems, or other NPL related tasks. After its normalization, an analysis of the Corpus is presented, as well as analyzes performed after the corpus vectorization (embedding) generated with the use of two current techniques.*

Keywords: *text generation; Corpus linguistics; lyrics*.

2.1.1. Introduction

Corpora are linguistic resources that are difficult to create and time-consuming. Nevertheless, they are very useful resources for language studies and training of natural language processing tools. In general, these resources are constructed from texts and digitized documents, in the case of corpora based on written natural language. In the case of corpora based on spoken natural language, conversations or interviews are used. More recently, the Web has presented itself as an important source of raw material for building corpora. There is a vast amount of textual and oral material in digital form and available in several languages. There is also a growing demand for large corpora due to the emergence of modern machine learning tools. As a result, Web-based corpora propositions are emerging. Habernal et al. (2016) presented a Multilingual Web-size *Corpus* containing over 10 billion tokens, licensed under Creative Commons license family in more than 50 languages. According to the authors, the texts that compose the *corpus* were extracted from CommonCrawl, the largest publicly available general Web crawl to date with about 2 billion crawled URLs. The size and diversity of the Web also allows the construction of large, specialized corpora. Seitner, Julian, et al. (2016) presented a publicly available database containing more than 400 million hypernymy relations extracted from the CommonCrawl web *corpus*. However, there are few corpora geared towards studies of poems and song lyrics.

In the present work we try to contribute to minimize this problem, presenting a song lyrics *Corpus* of music of random musical genres in English constructed with the use web scraping technique. The objective is to use the *corpus* as input for machine learning tools and automatic song lyrics generation. The article describes the process of extracting the song lyrics, the normalization phase, and the final analysis of the created *Corpus*. Next section presents *Corpus* development work that covers issues related to what is presented in this article. In section 2.1.3 is explained the web scraping method used to collect the music lyrics and how the data was processed after the collection, as well as the ‘‘cleanup’’ process adopted. In section 2.1.4, the *Corpus* analysis is presented, where N-grams, noisy words and other features are presented. In section 2.1.5, two different forms of embedding applied in *Corpus* are compared, and finally we present a brief conclusion along with the indication of future work in Section 2.1.6.

2.1.2. Related Works

Here we present which bear a certain resemblance to the current work. As mentioned earlier, there are not many works related to creating lyrics *corpus*, but it is possible to find some related datasets.

Nishina (2017) investigated various features identified in the lyrics of contemporary popular songs ranked in the Billboard Hot 100 chart, covering the 2002-2011 period. The author gathered from the site a total of 1,000 songs and, after that excluded noise characters such as leading whitespace. The resulted *corpus* presented an average of 502 tokens, and an average of 149 types for the 10-year period. Subsequently, the author performs a linguistic analysis on the *corpus*, analyzing the genre of the songs and the expressions used according to the sex of the author of the lyrics. The main difference of this work for the current work is the size of the *corpus*. Being such a small *corpus* is not suitable for quantitative analyzes and to feed machine learning tools, lending itself more for qualitative analysis.

Miethaner (2001) developed BLUR (Blues Lyrics) *corpus*, containing blues lyrics from the early twentieth century, focusing on the study of syntactic phenomena in earlier African American English. The *corpus* is composed of a computerized collection of more than 8,000 transcripts of pre-World War II blues recordings. Like the previous article, the *corpus* developed is too small to be handled by machine learning tools. Machine learning algorithms tend to work better on larger datasets, due to the bigger quantity of examples.

Ellis et al. (2014) presented the LyricFind *Corpus*, developed at the Sound & Music Computing laboratory at the National University of Singapore. The *corpus* consists of 275,905 distinct lyrics in bag-of-words format (67.6 million tokens). This is a *corpus* that is worth mentioning due to its volume, though, because it is presented in the form of bag-of-words, it is not suitable for use in training for text generation.

Kuznetsov (2019) has collected 57650 songs acquired from LyricsFreak through scraping. According to the author, he did some basic cleaning on the lyrics, removing non-English lyrics, extremely short and extremely long lyrics, and lyrics with non-ASCII symbols. Compared with current work, the number of song lyrics is less than half, although volume provided is enough to employ machine learning techniques.

2.1.3. Extraction of Lyrics and *Corpus* Cleaning

Following the same web scraping flow described by Milev et al. (2017), the creation of the *corpus* began in the selection of a source where it was possible to extract a sufficiently large number of song lyrics. Initially, the site chosen was Genius.com, due to the fact that it is one of the most used websites in the field and contains explanatory notes in some stanze of the songs, which could later be used to enrich *Corpus* content. The site provides an Application Programming Interface (API) for extracting data from songs (lyrics, artists, and other metadata). However, its use is limited so that a maximum daily amount of extractions of the lyrics is imposed. For this reason, it was decided not to use it. Thus, it was decided that an open site would be used, where the number of requests was unlimited. The website musica.com contained, up to October 2018, an amount of 979,972 registered song lyrics. The format of the URL and the website page layout enabled a simple extraction of its information, and by applying a small script written Python it was possible to extract 120,946 lyrics of songs from the site. The website indexes its song lyrics by a unique id, which made the main section of the Web scraping tool work with a simple for loop starting at the “lyric 1” and ending up after 120,946 iterations.

The goal was to extract only lyrics in English, however, some lyrics in other languages were also downloaded. Some songs written in Spanish, German, French and Italian were detected. For this reason, the cleaning of the *corpus* began in the extraction process itself. In the Python script, the langdetect native language library was used to identify the natural language used in a text. After extracting the song lyrics from the site, two files were generated: one with the *Corpus* itself, where each song was represented by a single index followed by the lyrics, and another file containing the song metadata: the artist and the title of the song.

As was said, the first phase of the cleaning process was the elimination of song lyrics in languages other than English. Still, some lyrics in English have expressions in other languages, mostly in Spanish, since many Latin artists who produce music for the American market mix the two languages. Several tracks of music were found that blended two languages, mostly Spanish and English. The song "Bailamos" by Enrique Iglesias has excerpts such as “... te quiero amor mio, bailamos, gonna live this night forever”. The decision was to keep these lyrics with this characteristic due to the fact that this is a specific characteristic of certain artists, and often, of specific genres such as pop music.

The second processing phase of cleaning the *Corpus* involved the structure of the song lyrics. Many of the extracted song lyrics contained markers indicating repetitions of the lyrics elements and types of elements such as chorus and verses. So, markers such as [Verse], [Chorus], [Repeat 2x], [Repeat 3x], among others have been removed. In this second phase, once again, a Python script was used to remove these markers. Markers with names of artists that was intended to clarify which artist was responsible for singing a certain part of the song have been removed as well. For example, in the song "Home Alone" by R. Kelly, the singer has the collaboration of another artist, Keith Murray. The piece of music in which Keith sings is demarcated by [Keith].

After finishing the cleaning, two new versions of the *corpus* were generated: in addition to the original *corpus*, a tokenized and lemmatized *corpus* was generated, in order to reduce the vocabulary size. The spaCy library tokenizer (Honnibal and Montani, 2017) was used and for the lemmatization, it was used the Python NLTK library (Bird et al. 2009), which has a built-in implemented lemmatizer

(WordNetLemmatizer). Thus, for instance, in the song "If I Die 2Nite" by Tupac Shakur, the sentence ``A coward dies a thousand deaths. A soldier dies but once´´ would become “A coward die a thousand death. A soldier die but once”.

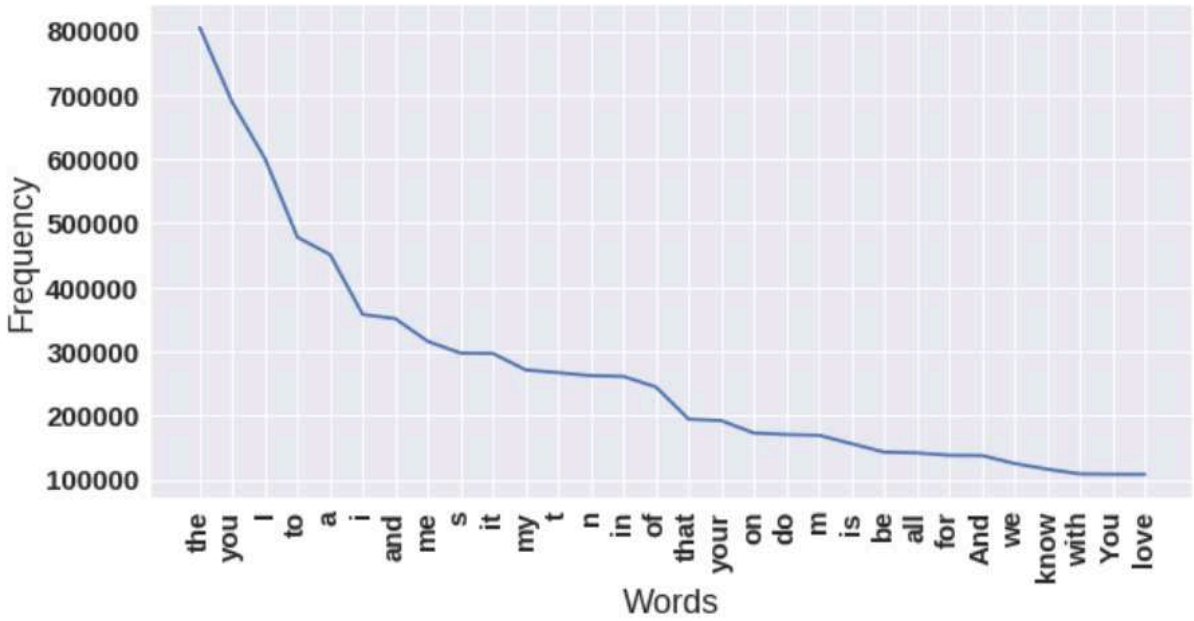
The second version of the *corpus* was generated when the stopwords were removed. Among the words present in the stopword list are “me”, "I", "myself", "you", "you're", for example. Note that these are words present in almost every song, and that often contribute to the expression of some feeling in the context. For example, in the song “Forever Man” by Eric Clapton, the phrase “How many times I say I love you” after processing would look like this “many times must say love”, which completely removes the sense of it. For this reason, the *corpus* version without stopwords was used only for N-grams analysis, described in the next section. Some other considerations were taken into account when analyzing the initial *corpus*. For example, the presence of onomatopoeias as “whoa”, “ooooh” in its most diverse forms, as well as the emphasis on certain syllables of some words to generate musicality, as in the case of “girl”, that several times was used like “Girrlll”. It was decided that such structures would be maintained because they were used to promote more musicality to the song and, in a way, highlight the given word in context. In addition, by maintaining such constructs, we avoid reducing the number of the *Corpus*’ types. The last consideration regarding the cleaning of *Corpus* was the removal of punctuation from both versions.

2.1.4. *Corpus* Analysis

Corpus analysis was done separately for the two versions described in the previous section. Firstly, an analysis of the frequency distribution was made in the two corpora, with the purpose of identifying the number of tokens and types in each one of them and to establish their size. In addition, an analysis of the occurrence of the unigrams in both was performed. For the lemmatized *Corpus*, 12,355,270 tokens and 175,412 types were counted. It is only natural that there is a much greater number of tokens than types, especially when it comes to song lyrics since there are many repetitions like what happens in choruses.

Afterwards, an analysis of the unigrams in the *Corpus* was carried out. In the figure below, the thirty most frequent words in the *corpus* are presented. As expected, the most common words in *Corpus* were words like “the”, “you”, “I”, “and”, among others.

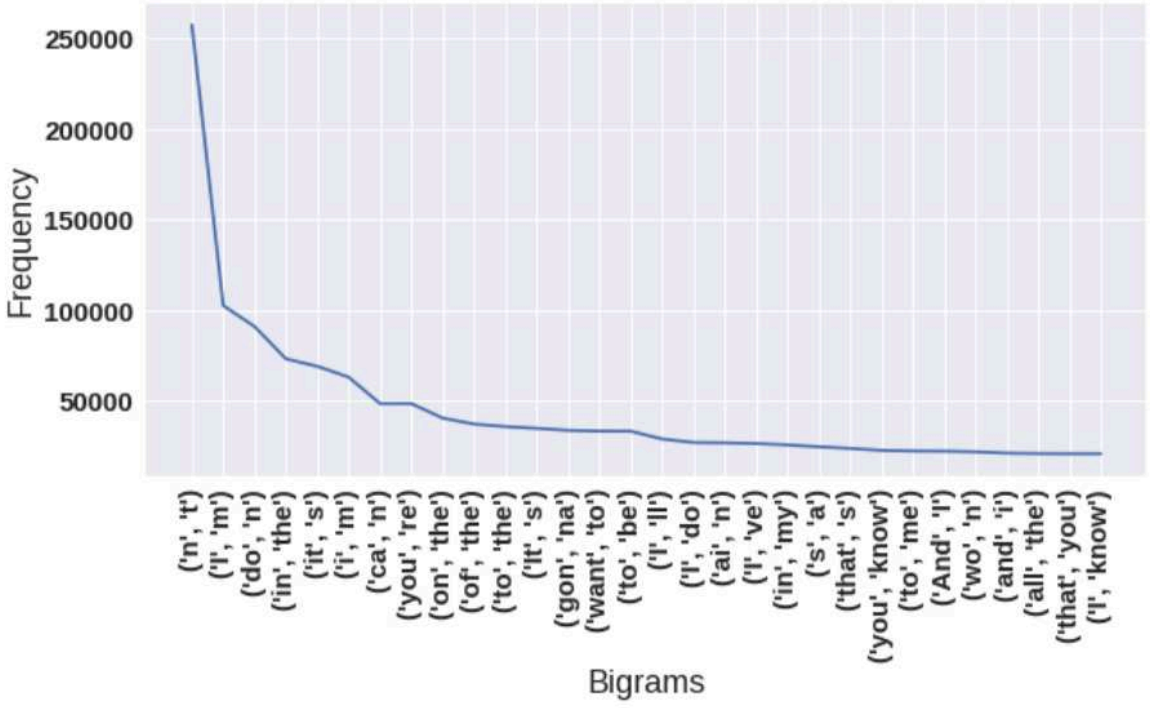
Figure 1 – The thirty most frequent unigrams in the lemmatized *corpus*.



Fonte: Elaborado pelo autor.

In the case of the *corpus* without stopwords, 11,300,686 tokens and 237,786 types were counted. The figure below shows the thirty most frequent bigrams in the *corpus*. Notably, the pronouns “you” and "I" participate in several bigrams among the most frequent.

Figure 2 – The thirty most frequent bigrams in the *corpus*.



Fonte: Elaborado pelo autor.

2.1.5. Embeddings

Two vectorization techniques were applied in the *corpus* and, following, an analysis was performed. First, the Word2Vec technique (Mikolov et al., 2013) was used through the Python Gensim package (Rehurek and Sojka, 2010). A similarity analysis was performed between words present in both versions of the *corpus* (lemmatized *corpus* and the *corpus* without stopwords).

Table 1 – Words similar to "man", "car", "love", "money", "drugs" and "life", with a window of words of size 5 in the lemmatized *corpus* using Word2Vec. These words were chosen randomly.

	woman	**guy**	**boy**	**girl**	**person**	**brother**
man	81.54%	73.77%	69.26%	68.97%	65.31%	63.31%
	truck	limousine	jeep	Cadillac	Benz	automo bile
car	79.21%	73.46%	72.41%	72.08%	70.80%	67.47%
	loving	promise	life	dream	trust	hope
love	69.47%	65.29%	64.18%	63.85%	63.35%	63.26%
	cash	dollar	dough	loot	cheese	respect
money	85.95%	85.61%	83.59%	74.13%	61.23%	61.18%
	dope	monopoly	cocaine	auto	junky	liquor
drugs	73.83%	60.12%	58.49%	58.22%	58.06%	57.22%
	world	existence	destiny	lifestyle	dream	fate
life	71.43%	68.75%	68.05%	66.20%	66.04%	65.98%

Fonte: Elaborado pelo autor.

Based on some of the results of the table above, it can be stated that Word2Vec was capable of capture the context of slangs. In America, people usually refer to the word money as cheese informally, and mainly in music. On the lemmatized lyrics *corpus*, the word cheese can refer to the noun cheese that represents food. As an example, the song “Summer Girls” composed by the artist LFO has the following use of the word cheese: “Think about that summer and I bug cuz I miss it. Like the color purple macaroni and cheese”. In the other hand, the group Cypress Hill use cheese as money on the song “Superstar”: “... come with me, show the sacrifice it takes to make the cheese”. The same analysis can be done to the word dope. During informal conversations it can be used as “cool” or “nice”. In this case, Word2Vec showed that the word dope was 73,83% similar to drugs, which is true, due to the fact that it is used on song lyrics to refer to illegal drugs taken for

recreational purposes. For the *corpus* without English stopwords, we have the following percentages of similarity.

Table 2 – Words similar to "man", "car", "love", "money", "drugs" and "live", with a window of words of size 5 in the *corpus* without stopwords using Word2Vec.

	woman	**guy**	**boy**	**girl**	**dude**	**kid**
man	71.75%	67.06%	65.43%	62.07%	57.16%	56.70%
	truck	cars	benz	bike	van	bus
car	72.61%	66.89%	62.54%	62.32%	61.52%	59.61%
	loving	know	oh	babe	baby	loves
love	66.68%	61.31%	60.58%	60.04%	59.86%	58.51%
	cash	dough	loot	moneys	funds	chips
money	79.58%	76.30%	63.59%	57.56%	56.74%	55.29%
	dealers	Drug	dope	dealer	cocaine	fiend
drugs	64.77%	62.09%	61.41%	59.57%	57.90%	56.40%
	world	lifes	lives	lifetime	love	time
life	66.32%	61.70%	61.27%	58.52%	55.51%	53.26%

Fonte: Elaborado pelo autor.

In order to perform a comparison, analyzes were also performed using another word-vectoring technique. The second vectorization technique used was fastText (Bojanowski et al., 2016). The great differential of fastText in relation to the representation made by Wor2Vec is that each word is represented as a bag of character n-grams in addition to the word itself. For example, be the word “money”, then using the n=4 parameter in the fastText configuration, it will generate representations for the character n-grams “”, using the characters “<” and “>” as boundary symbols. Because fastText constructs the vector for a word from n-gram vectors that constitute a word, it is able to output a vector for a word that is not in the pre-trained model. This can be quite interesting for lyrics generation since it works with phonetic similarity. Table 3 shows the words that are similar to "man", "car", "love", "money", "drugs" and "life", with a window of n-grams of size 3 in the lemmatized *corpus* using FastText.

Table 3 – Words similar to "man", "car", "love", "money", "drugs" and "life", with n=3 in the lemmatized *corpus* using FastText.

	woman	**catwoman**	**Woman**	**fellowman**	**mr.man**	**lawman**
man	77.9%	77.82%	77.34%	75.51%	75.2%	75.09%
	truck	limousine	Cadillac	houseboat	drivewa y	Carib
car	79.57%	76.86%	76.17%	76.03%	74.81%	74.41%
	loving	promise	life	dream	trust	hope
love	69.47%	65.29%	64.18%	63.85%	63.35%	63.26%
	cash	dollar	dough	loot	cheese	respect
money	85.95%	85.61%	83.59%	74.13%	61.23%	61.18%
	dope	monopoly	cocaine	auto	junky	liquor
drugs	73.83%	60.12%	58.49%	58.22%	58.06%	57.22%
	world	*existence*	*destiny*	*lifestyle*	*dream*	*fate*
life	71.43%	68.75%	68.05%	66.20%	66.04%	65.98%

Fonte: Elaborado pelo autor.

2.1.6. Conclusions

In the present article the development of a *corpus* containing song lyrics was presented. The *corpus* underwent a normalization process, and two versions were created, one where the words were placed in their lemma form and the other where the stopwords were taken. No syntactic or semantic annotation process has been done, leaving these stages as future work. The *corpus* is, so far as it has been found in current literature, the largest, except those that are in the form of bag-of-words, but which are not suitable for text generation tools. We pretend to make the *Corpus* available online on a website such as Kaggle. It is expected that the availability of the *corpus* allows the development, testing and evaluation of tools that seek the generation of text focused on poetry and song lyrics.

2.2. Geração de Letras de Música apoiada por Modelos Pré-treinados

Resumo. *Os avanços nas arquiteturas de redes neurais têm permitido melhorar a qualidade de diversas tarefas dentro do escopo da linguística computacional. Dentre*

as tarefas beneficiadas, podemos mencionar os sistemas de perguntas e respostas, sistemas de diálogos, mineração de opinião e a geração automática de textos, apenas para mencionar algumas. Apesar dos avanços, ainda existe espaço para contribuições, uma vez que existem problemas ainda em aberto. No caso da geração de texto, principalmente no gênero musical, existem desafios para a produção de textos que envolvem poesias e figuras de linguagem. Particularmente, alguns destes desafios estão relacionados com o tratamento de metáforas e metonímias e geração de paráfrases. O presente artigo apresenta uma análise de geração de trechos de letras de música tendo por base um modelo de rede neural GPT-2 pré treinado, após realizado o fine-tuning com dois corpora de letras de música, um em inglês e outro em português. É apresentada uma análise da grafia, sintaxe e semântica dos textos gerados, análise da coesão semântica dos mesmos a partir da métrica da perplexidade, assim como a discussão sobre a tentativa de encontrar um padrão nos trechos gerados pela ferramenta implementada. A pesquisa evidencia a possibilidade de uso de tais modelos na geração de textos poéticos, visto que a ferramenta desenvolvida foi capaz de gerar trechos textuais autorais e com medidas de perplexidade consideradas aceitáveis.

Palavras-chave: *Processamento de Linguagem Natural; geração de texto; letras de música.*

Abstract. *Advances in neural network architectures have allowed to improve the quality of several tasks within the scope of computational linguistics. Among the tasks benefited we can mention question and answer systems, dialogue systems, opinion mining and the automatic generation of texts, just to mention a few. Despite the advances, there is still room for contributions since there are still open problems. In the case of text generation, especially in the musical genre, there are challenges for the production of texts that involve poetry and idioms. In particular, some of these challenges are related to the treatment of metaphors and metonymies and the generation of paraphrases. This article presents an analysis of the generation of excerpts of lyrics based on a pre-trained GPT-2 neural network model, after fine-tuning with two lyrics corpora, one in English and one in Portuguese. An analysis of the spelling, syntax and semantics of the generated texts is presented, as well as the discussion about the attempt to find a pattern in the sections generated by the implemented tool. The research shows the possibility of using such models in the generation of poetic texts.*

Keywords: *Natural Language Processing; text generation; lyrics.*

2.2.1. Introdução

O Processamento da Linguagem Natural (PLN) apresentou grandes avanços nesta década, ao se beneficiar das recentes arquiteturas de redes neurais profundas (DENG e LIU, 2018). As tarefas dentro do escopo da PLN atingiram novos níveis no estado-da-arte, como foi o caso da tradução automática, da mineração de opinião, de sistemas de diálogo e de pergunta e resposta e de geração de texto. Hoje é possível adquirir, em lojas comerciais, assistentes pessoais que interagem por meio

de linguagem natural com uma certa fluidez, tais como Alexa, Siri e Google Assistant (HOY, 2018).

Dentre as tarefas relacionadas à PLN, uma das que tem atraído a atenção dos pesquisadores é a geração de linguagem natural (natural language generation - NLG). Em uma busca pelo termo “text generation” no Google Acadêmico, em janeiro de 2021, filtrando apenas artigos a partir de 2010, foram retornados 15200 links. Esta é uma tarefa atraente pois pode ser combinada com outras tarefas, gerando sistemas de diálogo e de pergunta/resposta mais naturais e com maior fluidez. Ela é também uma tarefa importante, isoladamente, tendo por objetivo produzir textos inéditos, com performance equivalente à dos seres humanos (GATT and KRAHMER, 2018). Como destacado por Gatt e Krahmer (2018), definir exatamente o que é NLG é mais difícil do que parece ser inicialmente. Apesar do produto final ser constituído de enunciados em linguagem natural, o que serviu como entrada para a produção do texto pode variar enormemente. Apenas para citar alguns exemplos, existem sistemas que produzem textos a partir de imagens (VINYALS et al., 2015), a partir de um único outro texto, como o caso de sumarização de texto (TAS and KIYANI, 2007), ou a partir de modelo neural previamente treinado com o uso de um *Corpus* linguístico (RADFORD et al. 2019). O presente trabalho se enquadra nesse último caso.

Mais recentemente, graças aos avanços no hardware e na arquitetura de redes neurais, foi possível o desenvolvimento de modelos neurais com bilhões de parâmetros, como é o caso do GPT-2 (RADFORD et al. 2019) e GPT-3 (BROWN et al., 2020), ambos desenvolvidos pela OpenAI. Estes modelos foram treinados com um alto custo computacional e, consequentemente, energético (STRUBELL; GANESH and MCCALLUM, 2019) e, por isso, precisam ser incorporados em outros sistemas, por meio de transferência de aprendizado (transfer learning), de modo a justificar o custo de seu treinamento e evitar maiores impactos ambientais.

Neste artigo, é verificado se é possível usar um desses modelos, mais especificamente a GPT-2, na tarefa de geração de linguagem natural de texto poético/letras de música. A escolha do modelo citado anteriormente se deu devido ao fato de, no momento em que se deu início ao trabalho em questão, GPT-2 ser o estado da arte dentre os modelos de linguagem. Optamos por mantê-lo mesmo que novos modelos, como GPT-3 tenham sido lançados durante a pesquisa. Neste artigo são discutidas as características do modelo implementado com base nos samples (exemplares de letras de canções, poesias) gerados a partir da execução do modelo.

O resultado é avaliado com respeito à estrutura textual, os padrões detectados nos textos gerados, proximidade dos samples com as letras de músicas compostas por seres humanos, dentre outros critérios. Foram implementados algoritmos geradores de samples textuais com utilização de modelos simplificados da GPT-2 fornecidos ao público pela OpenAI (https://openai.com/), sendo gerado um total de 10 amostras para posterior análise e discussão. O modelo foi treinado com *corpus* de letras de música em inglês elaborado por Rodrigues et al. (2019). Foi utilizado também um *corpus* de letras de música em português extraído do website Vagalume (https://www.vagalume.com.br/), para fins de análise da potencialidade da arquitetura na geração de letras de música na língua portuguesa.

O artigo está organizado da seguinte forma: na próxima seção discutimos os trabalhos relacionados com a presente pesquisa. Na seção 2.2.3 discutimos a tarefa de geração de texto em linguagem natural, seus principais desafios e estado atual. Na seção 2.2.4 apresentamos a abordagem adotada, bem como os recursos

empregados na pesquisa. Na seção 2.2.5 apresentamos os resultados obtidos e, finalmente, na seção 2.2.6 apresentamos as conclusões do trabalho.

2.2.2. Trabalhos Relacionados

Park e Ahn (2018) apresentam um modelo para geração automática de sentenças a partir de palavras-chave utilizando redes neurais recorrentes do tipo LSTM (Long Short-Term Memory) organizadas na forma de redes adversárias (Generative Adversarial Network - GAN). O modelo também inclui um módulo de self-attention. Os autores utilizam palavras-chave sinônimas como entrada do modelo a fim de aprimorar a quantidade de sentenças únicas geradas pelo mesmo. O modelo proposto pelos autores teve desempenho superior aos modelos que não usam GANs. Diferentemente do modelo proposto neste artigo, os autores não trataram de textos poéticos e não utilizaram modelos pré-treinados.

YI et al. (2018), abordaram dois problemas na geração automática de poesia, que é a falta de diversidade e o descasamento da avaliação de perdas, que são causadas por modelos neurais baseados em estimativa de máxima verossimilhança. Para lidar com esses problemas, eles utilizaram aprendizado por reforço e modelaram diretamente critérios e os utilizaram como recompensas explícitas para orientar a atualização do gradiente. O modelo foi baseado em redes neurais recorrentes do tipo GRU (Gated Recurrent Unit). Os autores trabalharam com poesia chinesa e, segundo eles, os resultados superaram o estado da arte. O diferencial em relação ao trabalho atual é o fato de não trabalharem com a língua inglesa e portuguesa e de não utilizarem modelos pré-treinados.

Van de Cruys (2020), propõe um modelo de redes neurais recorrentes do tipo GRU para geração de texto poético. O modelo foi treinado exclusivamente em texto padrão não poético, sendo usado para a geração de poemas em inglês e francês e, de acordo com os autores, o sistema produziu resultados compatíveis com o estado da arte para geração de poesia. Apesar de tratar de textos poéticos, as diferenças do trabalho com a pesquisa deste artigo são a não utilização de *Corpus* poético para treinamento e o fato de não utilizarem modelos pré-treinados.

2.2.3. Geração de linguagem natural e de Texto Poético

A partir de 2000, uma subárea do aprendizado de máquina, denominada de Aprendizado Profundo (do inglês Deep Learning), passou a se destacar, obtendo ótimos resultados na análise de dados não estruturados, tais como imagens e textos. Esses avanços foram alavancados pelo artigo de Hinton (HINTON, 2007) que mostrava como uma rede neural de retropropagação de múltiplas camadas poderia ser eficientemente pré-treinada, uma camada por vez, na forma de uma máquina não supervisionada restrita de Boltzmann, onde o objetivo é mesclar aprendizado supervisionado e não supervisionado num mesmo método. A partir de então uma ampla área de pesquisa foi gerada, apresentando resultados com precisão superiores aos métodos aplicados anteriormente. Dois fatores foram importantes para o crescimento acelerado da área: o desenvolvimento de placas gráficas capazes de realizar processamento paralelo com eficiência e baixo custo; e o

desenvolvimento de algoritmos de aprendizado eficientes para redes neurais com grande número de camadas, como o proposto por Hinton (HINTON, 2007).

Uma dessas áreas de pesquisa que floresceu foi o estudo de geração de textos a partir das estruturas gramaticais capturadas por esses novos métodos de aprendizado de máquina. De acordo com Piccialli et al. (2017), o campo de geração de linguagem natural consiste na criação de textos que fornecem informações contidas em outros tipos de fontes (dados numéricos, gráficos, taxonomias e ontologias ou mesmo outros textos), com o objetivo de tornar esses textos indistinguíveis, na medida do possível, daqueles criados por humanos. Para Fiorin e Savioli (1991), um texto se caracteriza por um composto de palavras ou informações que juntas fazem sentido, ou seja, é importante que um texto componha um todo que tenha algum significado e que possa ser interpretado por seres humanos. A geração automática de texto possibilita o aumento da produção de material textual que pode ter diversas finalidades, tais como, produção de material didático, produção de manuais técnicos, auxílio na produção de material de divulgação científica, geração automática de propaganda, etc.

Dentre os formatos de texto que mais estão presentes em nosso dia a dia, se destacam as letras de música e as poesias. Assim como qualquer outro tipo textual, é possível também obter resultados interessantes para a geração automática de conteúdo musical e poético utilizando ferramentas de PLN, o que será tratado no presente artigo. Artefatos da linguagem tais como as rimas, entonação, citações, dentre outras, faz com que o objetivo de gerar tais textos (baseando-se na definição de “texto” de Fiorin e Savioli) se torne um pouco mais distante.

2.2.4. Materiais e Métodos

O modelo para geração de letras de músicas proposto neste artigo se baseia no fine-tuning de um modelo pré-treinado, uma técnica denominada de transferência de aprendizado. Esta estratégia possui o benefício de se reaproveitar todo o custo computacional e energético investido no pré-treinamento. O modelo pré-treinado adotado foi o GPT-2 (RADFORD et al., 2019) fornecido ao público pela OpenAI (https://openai.com/). O GPT-2 é um modelo multitarefa pré-treinado que adota a arquitetura Transformer (Vaswani et al., 2017) e possui por volta de 1,5 bilhões de parâmetros. Por ser multitarefa, o modelo pode ser empregado em diversas tarefas de processamento de linguagem natural, tais como reconhecimento de entidades nomeadas, sistemas de pergunta e resposta, tradução, sumarização e geração de linguagem natural.

Já a arquitetura Transformer, proposta por Vaswani et al. (2017) no artigo intitulado “Attention Is All You Need”, é um modelo de arquitetura baseado unicamente nos mecanismos de atenção e que dispensa, totalmente, a utilização de redes recorrentes e redes convolucionais. Os autores afirmam que os modelos de redes neurais recorrentes, LSTM’s eram, na época da publicação, considerados estado da arte para diversas tarefas, tal como tradução automática (do inglês, machine translation). Entretanto, também é explicitado que os modelos recorrentes possuem uma natureza sequencial, de forma que uma sequência de estados ocultos h^t é gerado em função de um estado h^{t-1} anterior e uma entrada t. Por consequência dessa natureza sequencial, a paralelização dos exemplos de treinamento é prejudicada, tornando-se crítico para sequências maiores, dadas as restrições de

memória. Os autores ainda comentam que algumas melhorias foram apresentadas em trabalhos recentes, tal como alguns métodos de fatoração e computação condicional, entretanto, o problema com a natureza sequencial ainda persiste. Outro problema das redes recorrentes é atribuir um peso maior nas observações mais recentes e não conseguir reter informação quando a dependência entre os elementos é muito distante.

Tais problemas serviram de motivação para que fosse pensado no modelo Transformer, que passou a se basear num mecanismo de atenção, evitando a recorrência. Já no caso das redes convolucionais, a limitação está no fato que a percepção do contexto fica limitada ao tamanho dos filtros utilizados e que o aumento do tamanho dos filtros aumenta a complexidade do modelo, tornando-o difícil de treinar. Outro problema das redes convolucionais é o uso da camada de pooling que, quando empregada, causa perda de informação.

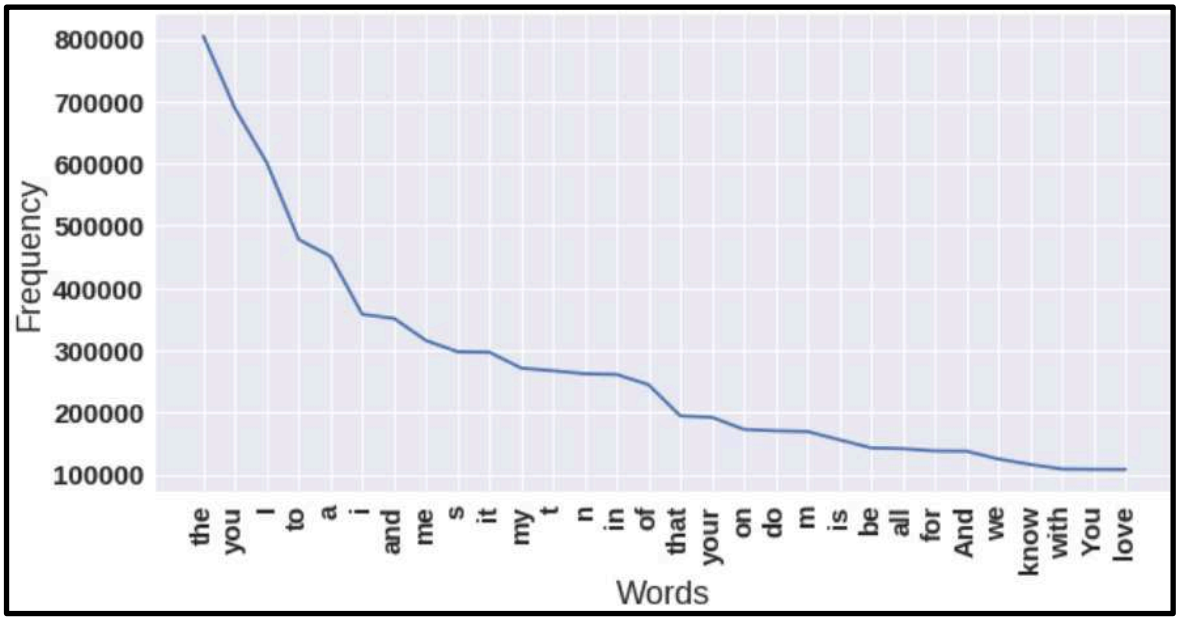
Além de adotar a arquitetura Transformer, o GPT-2 é adequado para ser empregado em transferência de aprendizado de modelos não supervisionados em uma configuração de Zero-shot. Em um aprendizado Zero-shot, o classificador é treinado com um conjunto de rótulos e é avaliado em um conjunto diferente de rótulos que o classificador não viu antes. No artigo de Radford et al. (2019), intitulado Language Models are Unsupervised Multitask Learners, o GPT-2 é empregado em diversas tarefas de PLN, sem a realização de fine-tuning e, ainda assim, alcança resultados de estado-da-arte em 7 de 8 dos conjuntos de dados testados. No trabalho aqui apresentado, o objetivo é utilizar o modelo GPT-2 a fim de comprovar se é possível gerar textos de cunho musical que estejam gramaticalmente corretos e que façam sentido como um todo. Tarefa para a qual o GPT-2 não foi treinado previamente.

Alguns modelos de GPT-2 foram disponibilizados para o público com intuito de testar a eficácia dos mesmos para uma gama de tarefas. O GPT-2 possui algumas versões de acordo com o número de parâmetros do modelo. Existe a versão com 117, de 345, de 762 e a de 1542 milhões de parâmetros. Quanto maior o número de parâmetros, maior o potencial de desempenho nas tarefas, no entanto, quanto maior o modelo, maior a exigência sob o hardware onde o modelo será executado. No caso da pesquisa descrita neste artigo, foram utilizados os modelos 345M e 762M para fins comparativos de potencial generativo dos modelos.

Para a realização do fine-tuning foram utilizados dois *Corpus* em duas línguas diferentes: inglês e português. Os corpora foram gerados por meio de Web Scraping em sites de música, sendo que o *corpus* em inglês (*corpus* 1) foi apresentado no artigo intitulado "Development of a Song Lyric *Corpus* for the English Language" (Rodrigues et al., 2019), e o *corpus* em português (*corpus* 2) foi criado a extração de letras de música do site Vagalume, por ser uma fonte reconhecida de letras de músicas nacionais. O processo de limpeza do *Corpus* 2 consistiu em sua simples tokenização e lematização, não houve uma limpeza rigorosa e minuciosa como foi realizado no *Corpus* em inglês.

Em relação ao *corpus* 1, após a lematização foram contabilizados 12.355.270 tokens e 175.412 tipos. É natural que ocorra um número muito maior de tokens do que de tipos, principalmente no que diz respeito às letras das músicas, pois existem muitas repetições como o que acontece nos refrões. Foi feita uma análise dos unigramas do *Corpus* 1, sendo que na figura 1, são apresentadas as trinta palavras mais frequentes do *corpus*. Sem surpresa as palavras mais comuns no *Corpus* foram palavras classificadas como artigos (the), pronomes (you, I, my, your), preposições (to, of) e conectivos (and).

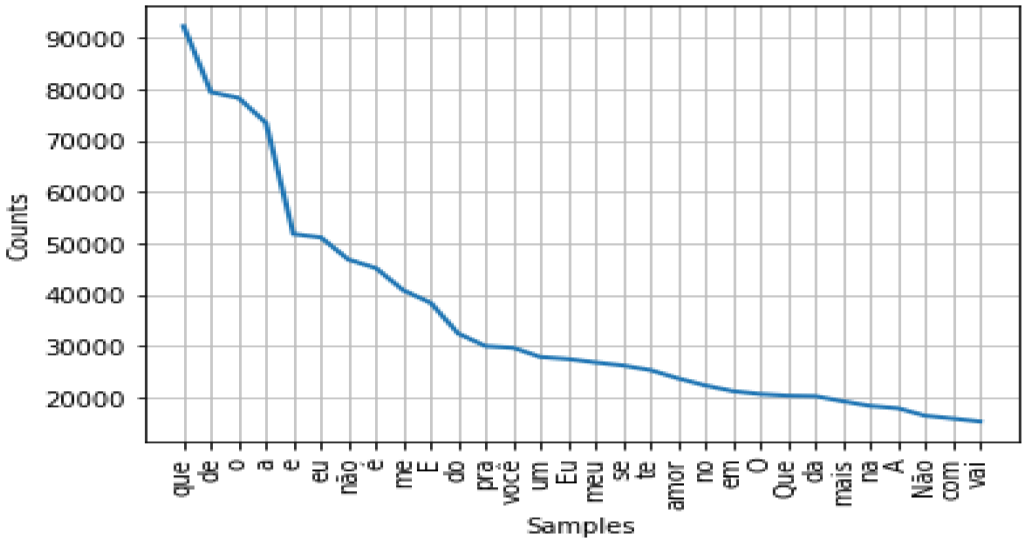
Figura 1 – As trinta palavras mais frequentes do *corpus* 1 após lematização.



Fonte: Rodrigues et al. (2019).

Em relação ao *corpus* 2, após a lematização foram contabilizados 3.761.958 tokens e 96.358 tipos em 24.783 canções. Na figura 2, são apresentadas as trinta palavras mais frequentes do *corpus*. Novamente, sem surpresa, as palavras mais comuns no *Corpus* foram palavras classificadas como artigos (o, a), pronomes (eu, você), preposições (pra, de) e conectivos (e, que). No *Corpus* com letras em português, algumas músicas possuíam trechos em espanhol, devido ao fato de alguns artistas brasileiros terem canções nesta língua, como por exemplo, a canção “Me Gusta”, da cantora Anitta. Não houve uma preocupação em remover tais canções, visto que, inicialmente, a frequência das mesmas foi considerada irrelevante para a tarefa.

Figura 2 – As trinta palavras mais frequentes do *corpus* 2 após lematização.



Fonte: Elaborado pelo autor

Os corpora foram usados para realizar o fine-tuning do modelo, porém, é necessário tomar alguns cuidados quanto à escolha do *Corpus* visto que, caso o seja pequeno e o fine-tuning seja executado por um tempo muito grande, é possível que ocorra o overfitting, que acontece quando o modelo desempenha bem no conjunto de dados de treinamento, mas não tem um bom desempenho com o conjunto de dados de teste. No caso da pesquisa realizada, não ocorreu esse problema, visto que, por exemplo, o *Corpus* 1 é composto por mais de 12 milhões de tokens.

Na próxima seção, são apresentados os resultados dos experimentos de fine-tuning usando os corpora. Os experimentos foram executados em uma máquina com 4 GPUs 2080 Ti e 116 GB de memória RAM para geração dos samples. Vale ressaltar que a GPT-2 aprende a predizer as palavras de acordo com o contexto, sendo assim, a performance não depende diretamente da língua, e sim da qualidade do conjunto de dados. Sobre a qualidade do dataset, entende-se pelo tamanho (medido em número de tokens e tipos) e da pipeline de pré-processamento do mesmo.

2.2.5. Resultados

Foram realizados quatro experimentos, resumidos na Tabela 1. Os experimentos foram divididos de acordo com o *Corpus* utilizado e com o tamanho do modelo. Os experimentos realizados como o modelo maior (762M) não puderam ser executados por muito tempo devido à limitação de hardware.

Tabela 2 – Experimentos realizados. Os experimentos foram separados em função do tamanho do modelo (milhões de parâmetros) e pelo *corpus* utilizado no fine-tuning.

Numero do experimento	Modelo	*Corpus*	Numero de amostras
1	345M	2	71
2	762M	2	18
3	345M	1	98
4	762M	1	6

Fonte: Elaborado pelo autor.

Para medir a evolução do texto gerado utilizamos uma métrica relacionada com a coesão semântica. Existem muitas propostas que podem ser usadas para medir a coesão semântica de um texto (NEWMAN et al, 2010). Nesta pesquisa foi utilizada a medida intrínseca perplexidade do texto gerado com base na probabilidade dos bigrams do *Corpus* original. A escolha dessa métrica baseou-se na facilidade de seu emprego e por ser fundamentada na probabilidade de coocorrência. Quanto menor o resultado do cálculo da perplexidade, maior a coesão semântica do trecho de texto. Não foi medida a qualidade da música gerada, uma vez que esse é um critério subjetivo. A fórmula de perplexidade utilizada está expressa na Equação 1.

$$Perplexidade = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(x_{i+1}|x_i)}} \qquad (1)$$

Onde: N é o número de tokens do texto analisado; x_i é o i-ésimo token do texto. Caso o bigram não ocorra no *corpus*, é utilizado a probabilidade do unigram multiplicado por um fator de suavização de 0,4.

Primeiramente, apresentamos o desempenho do modelo 345M no *corpus* em português. Foram gerados um total de 71 canções num período de 4:37h de execução do modelo. Nos samples gerados inicialmente, foi notado que o modelo ainda não era capaz de detectar algumas palavras, visto que, muitas delas não existem em nosso vocabulário. O Quadro 1 mostra um exemplo de saída com essa configuração do modelo.

Quadro 1 – Exemplo de saída do modelo 345M ajustado com o *corpus* 2, primeira amostra.

Vê se está, que a manera está,

que a fazer a sua vaz e a gato e eu se está,

que aqui tá quando o sua vaz se acabará;

Só tô quê ques más desprezú,

que a sua lado vai que viam para alguien sair!

Fonte: Elaborado pelo autor

É possível observar que é difícil estabelecer significado coerente no segmento apresentado no Quadro 1. Algumas palavras como “desprezú” e “ques” não sequer existem em português, e algumas palavras como “alguién” pertencem ao espanhol. No entanto, algumas destas palavras ocorrem no *corpus* original. Por exemplo, a palavra “desprezú” ocorre na música “Não há mais” de autoria de “Fm5”. Após algumas amostras geradas, já era possível notar evolução na coesão semântica, no que se refere à existência das palavras e colocação das mesmas. O contexto geral das canções já podia ser detectado (muitas das vezes o modelo usava o tema “amor”), porém, nota-se que algumas frases continham erros ortográficos ou eram incoerentes, como ilustrado no Quadro 2.

Quadro 2 – Exemplo de saída do modelo 345M ajustado com o *corpus* 2, segunda amostra.

No ar que não acontecer a esperança!

Se não ainda só você pra você!

Que eu perdi tanto eu não é pra mim!

Por que eu te quero você!

Você me fez apaixonar por aqui!

Eu tô olhar sem você!

Por que eu te quero você!

Eu tô olhar sem você!

Fonte: Elaborado pelo autor

No Quadro 3 é possível notar outra característica interessante dos textos gerados pela GPT-2 e a imitação da estrutura das músicas. Um grande número de letras de música apresenta elementos repetitivos, que são utilizados para prender a atenção, motivar e emocionar o ouvinte, seja através de um refrão, ou apenas uma

frase de impacto ao longo da canção. Observa-se na amostra apresentada no Quadro 3 a repetição das frases “Eu queria entregar” e “É pra te ver”.

Quadro 3 – Exemplo de saída do modelo 345M ajustado com o *corpus* 2, terceira amostra.

Deixa a gente viver
A mãe da cidade, vamos a bacada
Cidade de você, o meu chão da cidade
Eu queria entregar
Eu queria entregar
Eu queria entregar
Eu queria entregar tanto
Eu pude disse que não, no meu brado,
Não, não vai saber
Porque eu quis que eu
É pra te ver
É pra te ver
É pra te ver
É pra te ver

Fonte: Elaborado pelo autor

Ao final do treinamento, observa-se resultados interessantes na geração do texto. Alguns exemplares, inclusive, contam com marcações musicais, tais como a marcação de refrão e repetição. Além disso, notou-se que os assuntos abordados nas letras ficaram cada vez mais diversos. No Quadro 4 mostra o trecho de uma letra de música gospel que foi gerada. Em negrito, são destacadas as marcações.

Quadro 4 – Exemplo de saída do modelo 345M ajustado com o *corpus* 2, quarta amostra.

Que fica no amor, que fica no caminho de Deus (bis)
Estava solitário de tão profundo
Enche a vida como o tempo lá fora
E o melhor do Senhor

Há nenhum homem que estava no peito

E no fim de Jesus entrou na cruz

Que hoje o homem de Deus está perto

Santo, Santo, Santo, Santo ***(2x)***

[Refrão]

Amor mais belo e poderoso

És que eu vou chorar

Quem já sente por você?

Pois não há amor, sente só vocêr

Fonte: Elaborado pelo autor

No segundo experimento realizado, foi utilizado o mesmo *corpus* 2, porém, alterou-se o modelo para o 762M. Foram geradas 18 canções e, no geral, notou-se que os assuntos abordados nos exemplares gerados pelo modelo foram similares aos gerados no experimento anterior. Em sua maioria, assuntos como amor e religião tiveram maior destaque. Notou-se também que, mesmo que o modelo tenha sido executado por um tempo substancialmente menor devido à falta de recursos computacionais (o modelo 762M possui um número substancialmente maior de parâmetros, portanto, exige mais recursos). O exemplo do Quadro 5 foi gerado pelo modelo depois de uma hora de fine-tuning.

Quadro 5 – Exemplo de saída do modelo 762M ajustado com o *corpus* 2, primeira amostra.

Meus olhos não me deixam

Mas eu quero te ver que eu quero te amar

Cheguei de correr

Meus amigos não me deixam

Mas eu quero te ver que eu quero te amar

Cheguei de correr.

(REFRÃO)

Eu já sei que eu quero te ver

Você já sei que eu quero ter você

Eu já sei que eu quero te ver

Você já sei que eu quero ter você

Meus amigos não me deixam

Mas eu quero te ver que eu quero te amar

Cheguei de correr.

Fonte: Elaborado pelo autor

Na amostra do Quadro 5 é possível notar uma coesão semântica na letra gerada, ou seja, foi gerado um texto que faz sentido, apesar de que, por exemplo, a palavra “cheguei” em “cheguei de correr”, poderia ter sido substituída por palavras tais como “cansei”, “desisti”, “parei”, gerando um texto mais coerente. Outro ponto muito interessante, é o uso da conjunção adversativa “mas” de forma correta.

O exemplo do Quadro 6, mostra que o *corpus* tratou o assunto amor de uma forma negativa também, associando-o ao sofrimento que pode trazer. Acredita-se que o fine-tuning do conjunto de dados detectou esse viés negativo, que é comumente retratado nas letras da música popular brasileira.

Quadro 6 – Exemplo de saída do modelo 762M ajustado com o *corpus* 2, segunda amostra.

Se você quer ser você

Que não quer me ter, meu amor, pode sofrer

É assim, comigo eu estou

Eu queria amor

Eu queria o coração

Eu queria o infinito

Eu queria o coração

Não vou te mostrar você

Se você se esqueceu

Do que a gente viveu em mim

Eu queria o coração

Eu queria o infinito

Eu queria o coração

Eu queria o coração

Fonte: Elaborado pelo autor

Como já mencionado, o tema de maior destaque nas canções geradas foi “amor”. Isso pode ser analisado pela contagem das palavras no *Corpus* 2, uma vez que essa palavra ocorre com frequência. Tivemos resultados satisfatórios utilizando o *Corpus* em português como entrada do algoritmo, que por sua vez é substancialmente menor que o inglês. Para fins de comparação, o *corpus* em inglês foi composto por um total de 979.972 canções distintas e o conjunto de dados em português continha 24.783 canções. Logo, presumiu-se que os resultados em inglês seriam ainda mais interessantes e precisos.

Inicialmente, o modelo 345M foi testado para o inglês. Foram geradas 98 amostras num período de 9 horas. Logo no início da geração das amostras, já era possível perceber que a qualidade sintática e semântica superava os resultados anteriores. O Quadro 7 apresenta um trecho da primeira amostra gerada pelo modelo.

Quadro 7 – Exemplo de saída do modelo 345M ajustado com o *corpus* 1, primeira amostra

What an incredible moment of your life

You walked out of the movie theater

And into a rainstorm

On a moonlight night

Who knew you were the one you wanted to be

Oh the sky is blue

The moon is clear

You do n't remember me, but what I do know is

You were the star of the world

I knew the way you could be

You were the star of the star

I knew you were the star of the sun

Fonte: Elaborado pelo autor

Nota-se que o texto é sintaticamente correto (avaliado informalmente), no sentido de que não foi observado nenhum erro ortográfico, além do fato de que o texto gerado é um tanto quanto subjetivo, metafórico e poético. Alguns exemplos também mostraram que os textos não foram somente sobre amor e relacionamentos, mas também, aborda temas sexualmente apelativos e de poder, muito comum no gênero **rap**/**hip-hop.**

Quadro 8 – Exemplo de saída do modelo 345M ajustado com o *corpus* 1, segunda amostra.

Man 's in the house

Bury the child in the sand

My father made me a man

Bury the child in the sand

My father made me a man

He made me a man

I 'm a man

I 'm a man

I 'm a man with a gun

I 'm a man with a gun

I 'm a man with a gun

I 'm a man with a gun..

Fonte: Elaborado pelo autor

O exemplo do Quadro 8 fala de um homem com a posse de uma arma e que expressa através de sua canção que pode ser perigoso para a sociedade. Do ponto de vista sintático, notamos as frases repetidas (como dito anteriormente, comuns em músicas) e também a rima entre as palavras “sand” e “man”. Por sua vez, percebe-se que a GPT-2 não se mostrou eficiente no uso de rimas, dado que poucos exemplos continham essa característica. Alguns trechos estão em destaque no Quadro 9.

Quadro 9 – Exemplo de saída do modelo 345M ajustado com o *corpus* 1, terceira amostra.

*He 'll tell you he love **you***

*and he will tell you he feel the same way **too***

...

That’s an open wound, it is n't ***true***

It’s an open wound, I know all about ***you***

Fonte: Elaborado pelo autor

Infelizmente, por falta de recursos computacionais, não foi possível gerar muitos exemplares do *Corpus* em inglês para o modelo 762M. Sendo assim, foram geradas apenas 6 amostras. Mesmo assim, as amostras geradas se mostraram interessantes. O Quadro 10 mostra uma das saídas.

Quadro 10 – Exemplo de saída do modelo 762M ajustado com o *corpus* 1, primeira amostra.

I will always know you,

For all the things you love.

So sweet, so very very special!

This is my love.

My love I know you are in your heart.

My love, I know who you are.

I know where you come from.

And my love's only with me.

Don't let anyone hurt you.

Don't let me see you cry.

Don't let me make you sad.

Don't let me let you hide in your heart.

So sweet, very special, my love,

So very special.

Fonte: Elaborado pelo autor

É perceptível que algumas frases perdem o sentido, como a “Don't let me let you hide in your heart” ou a frase “My love I know you are in your heart”. A frase poderia ser substituída por “My love I know you are in [my] heart”.

A Tabela 2 mostra o resultado do cálculo de perplexidade de cada amostra gerada apresentada neste artigo. Estas amostras foram selecionadas por acreditarmos que são as mais representativas. É possível notar que existe uma tendência a diminuir a perplexidade à medida que o modelo vai sendo executado. No entanto, a mudança do tamanho do modelo parece não trazer benefício significativo. O melhor resultado com o *corpus* 2 foi obtido na amostra 5 com o modelo 762M, obtendo perplexidade 54. No entanto, o segundo melhor resultado com o *corpus* 2 foi obtido na amostra 2 com o modelo 345M, obtendo perplexidade 78. Já no caso do *corpus* 2, o melhor resultado foi obtido na amostra 8 com o modelo 345M, obtendo perplexidade 50. No entanto, os outros valores de perplexidade das outras amostras geradas com o *corpus* 2 não ficaram muito distantes deste valor. Em uma análise informal das amostras geradas, nota-se melhora nos resultados com o tempo de processamento e com o tamanho do modelo. Para se ter uma visão melhor das influências do tamanho do modelo e do tempo sobre os resultados seria preciso um número maior de amostras, no entanto, os custos de hardware para a execução destes modelos impediram a obtenção de um número maior de amostras.

Tabela 2 – Perplexidade das amostras do *Corpus* 1.

Amostra	Modelo	Perplexidade
7	345M	112
8	345M	50
9	345M	82
10	762M	74

Fonte: Elaborado pelo autor.

Tabela 3 – Perplexidade das amostras do *Corpus* 2.

Amostra	Modelo	Perplexidade
1	345M	1485
2	345M	78
3	345M	166
4	345M	138
5	762M	54
6	762M	102

Fonte: Elaborado pelo autor.

2.2.6. Conclusões

Todos os modelos de GPT-2 aqui analisados se mostraram capazes de gerar textos de cunho musical/poético de sintaticamente corretos e com coesão e semântica, mesmo com algumas limitações de recursos de hardware. Notou-se também que a ferramenta comete alguns erros ortográficos, que também ocorrem no *Corpus* original, porém, no geral, apresentou um resultado coeso, tanto sintática como semanticamente. Nesse sentido, a qualidade dos samples gerados se torna subjetiva aos olhos de quem os analisa, pois ainda não existem métricas capazes de analisar automaticamente a qualidade dos textos gerados segundo a perspectiva artística, sendo uma abordagem ainda aberta em NLP. Foi adotada a métrica de Perplexidade para avaliar os textos gerados, mas reconhecemos que essa métrica é insuficiente para medir a qualidade de um texto poético e serve apenas como ponto de partida para a análise deste tipo de texto.

Por fim, vale ressaltar que não é de interesse do projeto substituir completamente o esforço humano no que se diz respeito à composição musical, mas sim atuar como uma ferramenta de auxílio, como complemento do produto final. Não é de interesse provar que a GPT-2 pode substituir totalmente um artista humano, visto que, cada pessoa possui suas próprias formas de escrita, além do fato de que o teor pessoal trazido neste gênero textual se torna impossível, até o dado momento, de criar uma ferramenta que simule um compositor real, com sentimentos, vivências e crenças. Mesmo assim, acredita-se que a ferramenta tem um potencial único de geração textual e é genérico o suficiente para ser utilizado em outros gêneros textuais e com outros objetivos.

Como parte dos trabalhos futuros é preciso definir métricas mais apropriadas para tratar esse tipo de informação. É preciso também ampliar o número de experimentos a serem realizados, de forma a gerar resultados suficientes para uma análise mais aprofundada. Finalmente, outro trabalho futuro interessante seria a análise da incorporação nos conjuntos de dados de conteúdo semântico capaz de ajudar os modelos a produzirem textos poéticos de melhor qualidade.

Agradecimentos

Este estudo foi parcialmente financiado pela Coordenação de Pessoal de Nível Superior - Brasil (CAPES) - Código Financeiro 001, e também pelas agências de fomento FAPEMIG e CNPq.

3. CONCLUSÕES

Conclui-se com o presente trabalho que o desenvolvimento de *Corpus* textuais é uma tarefa de suma importância em NLP, visto que o mesmo serve como ponto de partida para demais tarefas, dentre elas, a geração automática de texto que também foi abordada no trabalho. As ferramentas para web scraping disponíveis atualmente, tal como a linguagem Python, juntamente com as bibliotecas de manipulação de páginas da Web possibilitaram a extração de um grande volume de dados, essencial para aplicações em NLP.

Concluiu-se também que, a utilização dos modelos de rede neural GPT-2 se mostrou capaz de cumprir a tarefa de automatic text generation, mesmo que os recursos computacionais utilizados não tenham sido diretamente proporcionais ao volume de dados e às exigências que tal tarefa impõe.

Infelizmente, ainda não se encontra na literatura um método exato para análise da qualidade dos exemplares de textos gerados pela ferramenta, entretanto, o uso da métrica de Perplexidade foi visto como forma de se medir a coesão semântica textual e os resultados se mostraram coerentes perante os diferentes modelos de GPT-2 e *Corpus* utilizados. Afirma-se que a ferramenta de geração textual utilizando GPT-2, previamente acompanhada da geração de Corpora com pouco ruído e suficientemente grande para ser utilizada como fine-tuning, pode atuar como forma de auxílio e fomentação da composição musical e poética.

Sendo assim, conclui-se que o trabalho aqui apresentado cumpre o que foi proposto inicialmente e propõe-se como trabalhos futuros o estudo e desenvolvimento de novas métricas para medição da qualidade semântica de exemplares textuais, o aumento no número de testes aliado a uma capacidade computacional mais robusta, o que também irá induzir a promoção de trabalhos relacionados utilizando Corpora mais robustas.

REFERÊNCIAS

BROWN, Tom. et al. Language models are few-shot learners. **34th Conference on Neural Information Processing Systems** (NeurIPS 2020), Vancouver, Canada, 2020.

BOJANOWSKI, Piotr, et al. Enriching word vectors with subword information. **Transactions of the Association for Computational Linguistics** 5, p. 135-146, 2017.

CARNEIRO DA SILVA, Jessica. **Da análise da música como gênero textual e texto multimodal ao ensino de língua portuguesa**. Trabalho de conclusão de curso em Licenciatura em Letras. UEFS, 2013.

DENG, Li; LIU, Yang. **Deep learning in natural language processing**. Springer, 2018.

ELLIS, R.J et al. Quantifying lexical novelty in song lyrics. **Proceedings of the 15th International Conference on Music Information Retrieval**, ISMIR 2014, Taipei, Taiwan,2014.

GATT, Albert; KRAHMER, Emiel. Survey of the state of the art in natural language generation: Core tasks, applications, and evaluation. **Journal of Artificial Intelligence Research**, v. 61, p. 65-170, 2018.

HABERNAL, Ivan; ZAYED, Omnia; GUREVYCH, Iryna. C4*Corpus*: Multilingual Web-size *Corpus* with Free License. **10th edition of the Language Resources and Evaluation Conference**, LREC, 2016, Portorož (Slovenia), 2016.

HINTON, Geoffrey. Learning multiple layers of representation. **Trends in cognitive sciences**, v. 11, n. 10, p. 428-434, 2007.

HOY, Matthew. Alexa, Siri, Cortana, and more: an introduction to voice assistants. **Medical reference services quarterly**, v. 37, n. 1, p. 81-88, 2018.

KUZNETSOV, S. **55000+ Song Lyrics**. [S. l.]. Disponível em: https://www.kaggle.com/mousehead/songlyrics. Acesso em: 9 jun. 2021.

MIETHANER, Ulrich. The BLUR (Blues Lyrics Collected at the University of Regensburg) *Corpus*: Blues Lyricism and the African American Literary Tradition. **Current Objectives of Postgraduate American Studies**. V. 2, 2001.

MIKOLOV, T. et al. Distributed representations of words and phrases and their compositionality. **In Advances in Neural Information Processing Systems**, p. 3111–3119, 2013.

MILEV, P. Conceptual Approach for Development of Web Scraping Application for Tracking Information. **Economic Alternatives**, Issue 3, p. 475-485, 2017.

NEWMAN, David et al. Automatic evaluation of topic coherence. In: **Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics**, p. 100-108, 2010.

NISHINA, Yasunori. A study of pop songs based on the billboard *corpus*. **International Journal of Language and Linguistics**, 4.2, 2017. p. 125-134.

PARK, Dongju; AHN, Chang. LSTM encoder-decoder with adversarial network for text generation from keyword. In: **International Conference on Bio-Inspired Computing: Theories and Applications**. Springer, Singapore, 2018. p. 388-396.

PICCIALLI, Francesco; MARULLI, Fiammetta; CHIANESE, Angelo. A novel approach for automatic text analysis and generation for the cultural heritage domain. **Multimedia Tools and Applications**, v. 76, n. 8, p. 10389-10406, 2017.

RADFORD, Alec et al. **Improving language understanding by generative pre-training**. [S. l.], 2018. Disponível em: https://openai.com/blog/language-unsupervised/. Acesso em: 9 jun. 2021.

RADFORD, Alec et al. **Language models are unsupervised multitask learners**. OpenAI blog, v. 1, n. 8, p. 9, 2019.

RODRIGUES, M.; OLIVEIRA, A. P.; MOREIRA, A. Development of a Song Lyric *Corpus* for the English Language. In: **NLDB - International Conference on Application of Natural Language to Information Systems**, 2019, Manchester. Natural Language Processing and Information Systems. Switzerland: Springer, 2019. v. 11608. p. 376-383.

SANTOS, Gilianne Vicente dos; LIMA, Alina Giseli da Silva; SILVA, Jacineide Virgínia Borges Oliveira. O uso do gênero letra de música para o desenvolvimento das competências linguístico discursivas dos alunos. In: **II Congresso Nacional de Educação**, Campina Grande, 2015.

SEITNER, Julian, et al. A Large DataBase of Hypernymy Relations Extracted from the Web. **10th edition of the Language Resources and Evaluation Conference**, LREC, 2016, Portorož (Slovenia), 2016.

STRUBELL, Emma; GANESH, Ananya; MCCALLUM, Andrew. Energy and Policy Considerations for Deep Learning in NLP. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. 2019. p. 3645-3650.

TAS, Guzman; KIYANI, Farzad. **A survey automatic text summarization**. Press Academia Procedia, v. 5, n. 1, p. 205-213, 2007.

VAN DE CRUYS, Tim. Automatic poetry generation from prosaic text. In: **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. 2020. p. 2471-2480.

VASWANI, A. et al. Attention is all you need. In **Advances in Neural Information Processing Systems**, 2017. p. 5998–6008.

VIEIRA, Armando. A Arte da Escrita Técnica. **Revista de Sistemas de Informação da FSMA**, V.8, 2011. p. 22-30.

VINYALS, Oriol et al. Show and tell: A neural image caption generator. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2015. p. 3156-3164.

YI, Xiaoyuan et al. Automatic poetry generation with mutual reinforcement learning. In: Proceedings of the 2018 **Conference on Empirical Methods in Natural Language Processing**. 2018. p. 3143-3153.