







PROPOE (prose to poetry): Geração Computacional de Poemas Metrificados a partir da Prosa Literária em Língua Portuguesa

PROPOE (prose to poetry): Computational Generation of Metrified Poems from Literary Prose in Portuguese Language

Ana Cleyge de Azevedo  
Universidade Estadual de Feira de Santana

João Queiroz  
Universidade Federal de Juiz de Fora

Angelo C. Loula  
Universidade Estadual de Feira de Santana

Resumo

A geração computacional do que pode ser concebido e reconhecido como ‘poesia verbal’ é explorada, há muitas décadas, em muitas línguas naturais. Mas apenas projetos recentemente desenvolvidos possuem aplicação intensiva da computação, considerando muitos níveis de organização, linguísticos e paralinguísticos, fonológicos, rítmicos, sintáticos, semânticos, e até pragmáticos. O que apresentamos aqui é um sistema de geração computacional de poemas, o PROPOE (*Prose to Poem*). Ele trabalha em conjunto com uma ferramenta de ‘mineração’ de estruturas de versificação na prosa de língua portuguesa, o MIVES (*Mining Verse Structure*). O PROPOE gera poemas em língua portuguesa a partir de sentenças versificadas (estruturas heterométricas de versificação) identificadas e classificadas pelo MIVES, e extraídas da prosa literária. O PROPOE combina sentenças, gerando poemas baseados na otimização de critérios rítmicos e fonológicos. É aplicado um ‘algoritmo guloso’ (*greedy algorithm*) cujo propósito é identificar a melhor combinação das sentenças, considerando normas rítmicas estabelecidas para o português. Em uma etapa final, realiza-se uma avaliação automatizada do resultado, atribuindo uma pontuação de acordo com a identificação de algum padrão considerado ótimo em poemas com métricas regulares, tendo como base esquema rítmico e adequação a estruturas rítmicas.

Palavras chave

geração de linguagem natural; poema; prosa literária; padrões rítmicos; rima

Abstract

The computational generation of what can be recognized as ‘verbal poetry’ has been explored for many decades in many natural languages. But only

recently developed projects have intensive application of computation, considering many levels of organization, linguistic and paralinguistic, phonological, rhythmic, syntactic, semantic, and even pragmatic. Here we present a computational poem generation system, PROPOE (Prose to Poem). It works in conjunction with a tool for ‘mining’ versification structures in Portuguese prose, MIVES (Mining Verse Structure). PROPOE generates poems in Portuguese from versified sentences (heterometric versification structures) identified and classified by MIVES, and extracted from literary prose. PROPOE combines sentences, generating poems based on the optimization of rhythmic and phonological criteria. A greedy algorithm is applied to identify the best combination of sentences, considering rhythmic norms established for Portuguese. In a final step, an automated evaluation of the result is carried out, assigning a score according to the identification of patterns considered optimal in poems with regular metrics, based on rhythmic scheme and adequacy to rhyme structures.

Keywords

natural language generation; poem; literary prose; rhythmic patterns; rhyme

1. Introdução

A geração computacional do que pode ser considerado ‘poesia verbal versificada’ é conhecida, ao menos, desde 1959, quando é publicado o ‘Stochastische Texte’ de Lutz (1959). No escopo mais amplo e histórico da ‘arte algorítmica’, ou ‘arte computacional’ (Antonio, 2009; D’Ambrosio, 2018), algoritmos ‘associa’ palavras para gerar versos (Lutz, 1959), caligramas ou estruturas verbo-visuais (Souza, 1991; Antonio, 2009), poemas baseados em sentenças extraídas de notícias de jornais, crônicas e de outros poemas (Balestrini, 1962; Antonio, 2009).



DOI: 10.21814/lm.14.2.369

This work is Licensed under a

Creative Commons Attribution 4.0 License

Em ciência da computação, o tema tornou-se um tópico de interesse no domínio da Geração de Linguagem Natural (Gatt & Krahmer, 2018) e Criatividade Computacional (Boden, 2009; Colton & Wiggins, 2012). Trabalhos pioneiros, baseados no uso intensivo de técnicas computacionais, foram propostos por Gervás (2000), Manurung et al. (2000) e Gonçalo Oliveira et al. (2007). Na última década, observamos um notável crescimento no número de trabalhos (ver revisão de Gonçalo Oliveira 2017b), uso de diversas técnicas, e o desenvolvimento de sistemas capazes de gerar poemas de variadas morfologias, atentos a diversos níveis de descrição – sintáticos, fonéticos, rítmicos, semânticos e até pragmáticos.

O desenvolvimento de um gerador de poemas consiste tanto na escolha geral de certas abordagens, técnicas associadas à eficiência e desempenho de processamento, quanto na construção de algoritmos capazes de seguir regras de combinação frásica, padrões de organização e versificação, e verificação do significado das palavras. Muitas vezes são eleitos apenas alguns níveis de descrição – sintático, gramatical, fonológico ou morfológico (Gonçalo Oliveira, 2017b). A reutilização, e combinação, de materiais previamente disponíveis (parágrafos, sentenças, versos), para geração de poemas, é uma estratégia usada por muitos pesquisadores (Gonçalo Oliveira, 2017b). Há muitos antecedentes na história recente da poesia do último século, em muitos idiomas (Perloff, 2013), e ela pode ser descrita como uma abordagem de ‘criatividade combinatorial’, como define Boden (2009), uma ‘produção de combinações não familiares de ideias familiares’.

Apresentamos um sistema de geração computacional de poemas versificados (PROPOE, *Prose to Poem*). Ele trabalha em conjunto com uma ferramenta computacional de mineração de estruturas de versificação, ou sentenças metrificadas, na prosa de língua portuguesa, MIVES (*Mining Verse Structure*)¹ (Carvalho et al., 2020). O PROPOE então seleciona e combina sentenças metrificadas, ou excertos de sentenças metrificadas, extraídas da prosa literária. Este fenômeno — ocorrência de estruturas de versificação na prosa literária, em língua portuguesa — foi identificado por Guilherme de Almeida (1946) em *Os Sertões*, de Euclides da Cunha. Almeida (1946) escandiu diversas sentenças metrificadas na obra de Euclides, e identificou uma grande variedade de estruturas, com ênfase em decassílabos he-

roicos e sáficos², e dodecassílabos³. Anos depois, Augusto de Campos (1996) produziu uma lista mais ampla, relacionou a operação com um método que chamou de ‘leitura verso-espectral’, e identificou ‘mais de 500 decassílabos significativos, com predominância de sáficos (acentuados na quarta e oitava sílabas) e heróicos (acentuados na sexta sílaba)’, e pouco mais de duas centenas de dodecassílabos, entre os quais muitos alexandrinos perfeitos. Para Almeida (1946), ‘se certas passagens d’Os Sertões, em vez de compostas tipograficamente em forma de prosa, o fossem em forma de versos livres’ seria possível encontrar poemas ‘autênticos’ e ‘legítimos’. Ele próprio, Almeida (1946), ‘gerou’ um poema versificado, ‘A Vaquejada’, através da redistribuição gráfica de um trecho:

A vaquejada

De repente estruge ao lado
um estrídulo tropel de cascos sobre pedras,
um estrépito de galhos estalando,
um estalar de chifres embatendo;
tufa nos ares, em novelos,
uma nuvem de pó;
rompe, a súbitas, na clareira,
embolada, uma ponta de gado;
e, logo após,
sobre o cavalo que estaca esbarrado,
o vaqueiro, teso nos estribos...

(Almeida, 1946, p.212-213)

As estruturas métricas definem a medida do verso (Spina, 2003, p.29), que pode ter diversas extensões. Em português, o sistema de estruturas de versificação é silábico-acental — conta-se o número de sílabas poéticas de cada verso e verifica-se a alternância entre sílabas fortes (tônicas) e fracas, até a última tônica. Este processo é chamado de escansão. Tal alternância estabelece um certo número de padrões que, combinados às repetições posicionais das sílabas poéticas, cria segmentos internos, estabelecendo as regras de versificação ou metrificação (Ali, 1999). O MIVES realiza a escansão de segmentos frásicos na prosa, separa as sílabas poéticas e classifica as estruturas encontradas de acordo com padrões rítmicos normatizados. Estas sentenças (metrificadas) fornecem material para o sistema PROPOE gerar os poemas, considerando aspectos rítmicos e fonológicos. O PROPOE usa um ‘algoritmo guloso’ para identificação de semelhanças entre padrões rítmicos. Explora di-

²Decassílabos heroicos e sáficos, correspondem a versos com dez sílabas poéticas cujas tônicas são identificadas nas posições 3, 6, 10, e 4, 8, 10 respectivamente (Spina, 2003)

³Dodecassílabos são versos compostos por doze sílabas poéticas (Spina, 2003)

¹<https://mivestool.wordpress.com/>

versos tipos de rima, considera várias estruturas métricas, combinando diferentes formas de busca através de similaridade entre versos, padrões rítmicos e posicionamento de tônicas. Baseado em critérios estabelecidos para avaliação dos poemas, o sistema avalia padrões fonéticos, rítmicos e similaridade entre as sílabas tônicas dos versos.

Nas próximas seções, apresentamos diversos projetos de geração automática de poemas, com ênfase em trabalhos considerados pioneiros, aplicações em vários idiomas, e destacamos alguns sistemas baseados na reutilização de textos. Em seguida, detalhamos o PROPOE, suas etapas de processamento, incluindo filtragem, montagem e avaliação dos poemas gerados. Finalmente, exibimos resultados de alguns poemas gerados pelo sistema, a partir de sentenças extraídas da prosa literária brasileira. Nossos resultados sugerem que há variações nas pontuações de avaliação obtidas pelos poemas, com possível relação com o volume de sentenças disponíveis, e que os diferentes critérios individuais de avaliação impõem níveis distintos de dificuldade para pontuação na avaliação.

2. Trabalhos Relacionados

Como mencionamos, são pioneiros os trabalhos de geração computacional de poemas desenvolvidos por Lutz (1959). Ele concebeu um sistema incapaz de combinar trechos de sentenças por meio de um algoritmo gerador de números aleatórios (Ryan et al., 2014). Mas este tópico tornou-se realmente científico, para a comunidade de Ciência da Computação, apenas a partir dos anos 2000, com tratamentos mais rigorosamente sistemáticos de muitos níveis, linguísticos e paralinguísticos (Gonçalo Oliveira et al., 2007). Os trabalhos de Manurung et al. (2000) e de Gervás (2000) deram início a novos projetos computacionais, muitos dos quais ganharam relevância através da aplicação de diversas técnicas de inteligência artificial.

Entre os projetos considerados mais promissores, e pioneiros, Manurung et al. (2000) aplicaram o modelo estocástico de *hill climbing*⁴ para implementar um sistema capaz de gerar poemas em inglês, a partir de buscas estocásticas por sentenças derivadas de regras gramaticais, lexicais e proposições semânticas. Em outra implementação, Manurung (2004) e Rahman & Manurung (2011) desenvolveram um sistema gerador

de poemas em inglês, o Mcgonagall (exemplo (a) na Tabela 1), que aplica algoritmos evolutivos⁵ para avaliação de padrões de tônicas e similaridade (métrica e semântica), utilizando operadores genéticos para variar características gramaticais, semânticas e lexicais de sentenças candidatas, representadas por árvores de derivação inicializadas aleatoriamente. Rashel & Manurung (2014) desenvolveram um sistema, o Pemuisi, que gera poemas em indonésio utilizando uma abordagem baseada na ‘satisfação de restrições’ (*constraint satisfaction approach*)⁶, tais como número de versos, contagem de sílabas e rima, a partir de notícias publicadas em jornais, e utiliza sentenças construídas a partir de modelos e palavras chaves obtidas nas notícias. Tobing & Manurung (2015) propuseram um sistema que não utiliza modelos de sentença, mas extrai sentenças de notícias online e aplica um ‘gerador tabular’ (*chart generation*)⁷ para produzir variações que atendam a um padrão determinado (exemplo (b) na Tabela 1).

O sistema WASP (Gervás, 2000) usa uma abordagem baseada em regras de ‘raciocínio avançado’ (*forward reasoning*)⁸ para produzir versos em espanhol a partir de um conjunto de palavras e padrões de versificação. Aplica-se um método para gerar e testar rascunhos de versos através de diversas restrições, produzindo versos independentes, ou poemas na forma de *romances*, *cuartetos*, e *tercetos encadenados*. Gervás (2001) incrementou sua implementação no sistema ASPERA (exemplo (c) na Tabela 1). O sistema solicita do usuário um fragmento de texto para referência, alguns parâmetros, seleciona métrica e estrofe em uma base de conhecimento, e aplica um método de ‘raciocínio baseado em casos’ (*Case-Base Reasoning, CBR*)⁹. Em outro momento, Gervás (2016) retomou seu projeto WASP para adicionar restrições semânticas (exemplo (d) na Tabela 1). Um *corpus* de poemas mexicanos, e notícias retiradas de jornais, foram utilizados para extrair sequências de palavras,

⁵Algoritmos Evolutivos são algoritmos de busca e otimização de uma população de soluções que passa por um processo iterativo de seleção competitiva seguida de variação (Eiben & Smith, 2003).

⁶Abordagens de ‘satisfação de restrições’ buscam soluções para atribuir valores para variáveis satisfazendo restrições que limitam seus possíveis valores (Brailsford et al., 1999).

⁷‘Geração tabular’ (chart generation) é um método para construir sentenças sintaticamente bem formadas de um enunciado de linguagem natural (Kay, 1996).

⁸‘Encadeamento para frente’ é uma forma de acionamento de regras em mecanismos de inferência (Durkin, 1994).

⁹‘Raciocínio baseado em casos’ (case-based reasoning) é uma técnica que resolve problemas buscando casos similares anteriores (Aamodt & Plaza, 1994).

⁴Subida da Encosta (Hill Climbing) é um método de busca local e iterativa de soluções para problemas de maximização, com avaliação e escolha pelos melhores estados com base no estado atual (Castro, 2006).

A very african lion,
 who is african, dwells in a waste.
 Its head, that is big,
 is very big.
 A waist, that is its waist, it is small.
 (a) (Manurung, 2004)

Ask in french surface
 Call her years, check her
 Think were toy tennis
 Skid in chase, land her
 (b) (Tobing & Manurung, 2015)

Ladrará la verdad el viento airado
 en tal corazón por una planta dulce
 al arbusto que volais mudo o helado.
 (c) (Gervás, 2001)

Toda era una ave larga
 que cuando se conforman.
 Admitió que se tienen
 registradas personas
 (d) (Gervás, 2016)

delatar sempre causa delação
 delação negra sem acusação
 acusação em meia delação
 sem achar cita, nem citação
 (e) (Gonçalo Oliveira, 2017a)

Tabela 1: Exemplos de poemas gerados por diferentes sistemas.

que são modificadas segundo critérios específicos, produzindo fragmentos de poemas, combinados em estrofes.

PoeTryMe é uma plataforma com arquitetura modular para geração de poemas em português desenvolvido por Gonçalo Oliveira (2012) (ver exemplos na Tabela 2). A plataforma disponibiliza várias estratégias de geração, incluindo sentenças aleatórias, sentenças selecionadas por método ‘gerar-e-testar’ (*generate-and-test*)¹⁰, ou sentenças obtidas por algoritmos evolutivos. Baseado em ‘palavras semente’ (*seed words*)¹¹, e utilizando redes semânticas construídas por dicionários, são gerados poemas conforme modelos (quantidade de estrofes, quantidade de linhas por estrofe e métrica) selecionados pelo usuário. Esta mesma plataforma foi utilizada por Gonçalo Oliveira (2017a) para gerar poemas baseados em postagens do Twitter, resultando no sistema *Poeta Artificial 2.0* (ver exemplo na Tabela 1e). O

¹⁰Em uma heurística de ‘gerar-e-testar’, soluções são produzidas de forma estocástica e então avaliadas por critérios específicos do problema.

¹¹Palavras de entrada inicial no sistema *PoeTryMe*, que servem como tema para geração do poema (Gonçalo Oliveira, 2012).

Poeta 2.0 reutiliza fragmentos de sentenças de postagens, com possibilidade de troca de palavras por sinônimos para satisfazer certos padrões métricos.

Haiku

ah paixão afecto
 não tem paixão nem objecto
 sem na arte modos

(palavras semente: paixão, arte)

Quadra

ah escultura representação
 e os que dão ao diabo o movimento da convulsão
 sua composição de preparação
 é destino estar preso por orientação

(palavras semente: paixão, arte)

Soneto

e não há deus nem preceito nem arte
 um palácio de arte e plástica
 as máquinas pasmadas de aparelhos
 num mundo de poesias e versos

o seu macaco era duas máquinas
 horaciano antes dos poetas
 para as consolas dos computadores
 num mundo de poesias e carmes

longas artes apografias cheias
 tenho poesias como a harpa
 poema em arte modelação

somos artificiais teatrais
 máquinas engenhocas repetido
 um poeta de líricas doiradas

(palavras semente: computador,
 máquina, poeta, poesia, arte,
 criatividade, inteligência,
 artificial)

Tabela 2: Exemplos de poemas gerados pelo *PoeTryMe*(Gonçalo Oliveira, 2012).

Outros pesquisadores desenvolveram geradores em diferentes línguas, baseados em diversas escolhas, relacionadas ao tratamento de diferentes fontes de dados, processamento e dados de saída. O sistema de Agirrezabal et al. (2013) gera poemas em basco (*euskera*), de tradição *bertsolaritza*, utilizando modelos de etiquetagem morfosintática (*POS tagging*)¹² obtidos de corpora de poemas. Das & Gambáck (2014) desenvolveram

¹²‘Etiquetagem morfosintática’ corresponde à atribuição de classes morfosintáticas às palavras de uma sentença, conforme sua contextualização. Para isso, utiliza-se tanto o léxico de palavras, e processos, categorizados pelo conjunto de etiquetas de um sistema, quanto o conjunto de possibilidades semânticas para categorias presentes no

um sistema de geração de poemas em bengali com sequências rítmicas determinadas a partir de uma linha de entrada fornecida pelo usuário, modelo da linha inicial e geração de linhas compatíveis a partir de um conjunto de poemas. A utilização de materiais previamente produzidos também é uma abordagem bastante conhecida. Tobing & Manurung (2015) e Gonçalo Oliveira (2017a) usam fragmentos de sentenças, modificados através de substituições de palavras. Outros projetos não modificam as sentenças, ou trechos de sentenças originais, e selecionam versos baseados em repositórios de sentenças. A partir de textos de blogs, o sistema de Wong et al. (2008) cria um acervo de fragmentos de sentenças com quatro palavras. Buscas por palavras-chave são realizadas neste acervo para obter haicais¹³, combinando sentenças através da similaridade de palavras. Netzer et al. (2009) cria um repositório de sequências de palavras obtidas de uma base do Google e do Projeto Gutenberg para construir haicais. Deste repositório são extraídas sentenças candidatas que satisfazem estruturas sintáticas, gerando haicais. Eles são, em seguida, submetidos a um ranqueamento pela presença de associação entre palavras para seleção de um poema final.

Além das abordagens citadas, pesquisas sobre a utilização de redes neurais artificiais, particularmente redes profundas, para geração de linguagem natural, têm aumentado nos últimos anos, embora a geração de poemas através dessa abordagem tenha sido indicada como a tarefa mais difícil por exigir a consideração de requisitos estruturais (Otter et al., 2020). Uma rede neural treinada com sequências de sílabas da *Divina Comédia*, de Dante Alighieri, foi utilizada por Zugarini et al. (2019) para gerar poemas em italiano em tercetos a partir de uma sequência inicial de sílabas aleatórias. O modelo pré-treinado da GPT-2 (Radford et al., 2019), treinado em 40GB de textos da Internet, também já foi aplicado para geração de poemas. Bena & Kalita (2019) utilizam este modelo com 774 milhões de parâmetros para gerar poemas em inglês, embora sem considerar requisitos estruturais. Hämäläinen et al. (2022) utilizou uma rede

GPT-2, treinada com 60GB de textos em belga, associada a uma segunda rede neural codificadora, para produzir poemas em francês, mas para obter rimas foi necessário utilizar um algoritmo que escolhesse uma palavra entre um conjunto de possibilidades indicadas pela rede GPT-2.

Há, portanto, uma grande variedade de sistemas de geração de poemas. Quanto às técnicas aplicadas, encontramos heurísticas como *hill climbing* e algoritmos evolutivos, sistemas baseados em regras, raciocínio baseado em casos, geradores tabulares, redes neurais, entre outras. Em relação aos níveis linguísticos, e paralinguísticos, de organização, são considerados aspectos gramaticais, morfológicos, fonéticos, rítmicos, semânticos, e até pragmáticos. Em geral, eles não são tratados conjuntamente, variando conforme a ênfase da pesquisa. O aproveitamento de materiais para início do processo de geração é uma estratégia frequente, como ‘palavras sementes’ (*seed words*) (Gonçalo Oliveira, 2017a), sentenças que sofrem ajustes, ou são preservadas (fragmentos) como versos.

O PROPOE utiliza sentenças extraídas da prosa literária. Isso é particularmente importante, no contexto de geradores automáticos, se considerarmos que a prosa literária ‘ocupa um lugar intermediário entre a poesia enquanto tal e a língua de comunicação comum, prática’ (Jakobson & Pomorska, 1985, p.106). Trata-se, também, de uma estratégia baseada na ideia de reutilização de paralelismos estruturais (fonológicos, rítmicos) já realizados, muitos dos quais já historicamente canonizados no acervo da fortuna crítica-literária em língua portuguesa. Nossas experimentações iniciais incluem a prosa de Euclides da Cunha, e dos modernistas Mário e Oswald de Andrade.

A reutilização de estruturas de uma fonte, sem modificação ou reposicionamento de palavras, aproxima o PROPOE de sistemas propostos por Netzer et al. (2009) ou Wong et al. (2008). Entretanto, estes dois sistemas anteriores reutilizam sequências curtas, de poucas palavras. O PROPOE é capaz de reutilizar sentenças inteiras, trechos iniciais ou finais de sentenças. Por outro lado, a construção de um poema é tratado pelo PROPOE como um problema de otimização, com foco em restrições e critérios rítmicos, a ser resolvido pela maximização de critérios definidos considerando restrições iniciais, como Manurung (2004) e Rahman & Manurung (2011).

léxico (Carvalho et al., 2012)

¹³Poema de origem japonesa, o haikai descende de um processo de adaptação cultural milenar que resultou numa vasta produção criativa (Sousa, 2007). Inicialmente chamado de waka (wa = expressão designativa de Japão e ka = poema, canto), aquilo que vem a se tornar o haikai é registrado pela primeira vez numa compilação de cerca de quatro mil e quinhentos poemas, distribuídos em vinte volumes, no primeiro e mais importante documento-livro poético da história do país, o Man’yōshū, ou *Antologia de dez mil folhas*, elaborado entre os séculos VII e VIII.

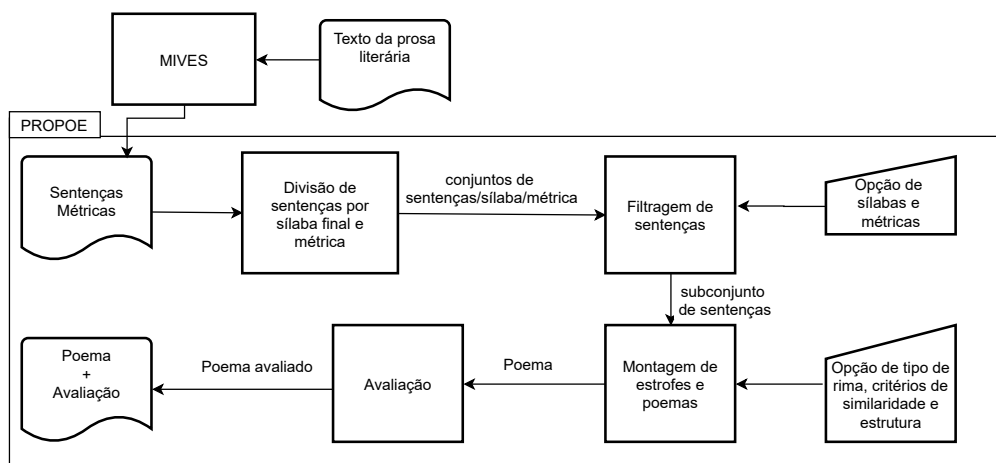


Figura 1: Etapas de processamento do sistema PROPOE.

3. PROPOE

O PROPOE¹⁴ (*Prose to Poem*) gera poemas versificados a partir de sentenças heterométricas (estruturas versificadas) identificadas e classificadas pelo MIVES (ver descrição abaixo). Estas estruturas são identificadas, classificadas e extraídas da prosa literária de língua portuguesa. O sistema combina tais estruturas e gera poemas baseados na otimização de critérios rítmicos e fonológicos. É aplicado um ‘algoritmo guloso’ (*greedy algorithm*) em que são consideradas normas rítmicas e rítmicas estabelecidas para o português (Silva, 2014). Em uma etapa final, é realizada uma avaliação automatizada, com atribuição de uma pontuação baseada em critérios relacionados a padrões rítmicos.

A arquitetura do sistema, e suas etapas de processamento, são apresentadas abaixo (Figura 1). Após o processamento pelo MIVES de um texto em prosa, as sentenças versificadas são enviadas ao PROPOE, que realiza a divisão e organização destas estruturas, indexadas de acordo com a última sílaba da última palavra da sentença e sua estrutura métrica. A partir das escolhas de sílabas finais e da estrutura métrica, é obtido um subconjunto de sentenças filtradas. Em seguida, são escolhidos os parâmetros estruturais que regulam a montagem das estrofes. De acordo com os critérios estabelecidos para avaliação, são exibidos os resultados. Detalhamos, em seguida, cada uma das etapas.

3.1. Prosa baseada em estruturas de versificação

Um texto em prosa de língua portuguesa é fornecido ao MIVES (Carvalho et al., 2020) que realiza a segmentação e o processamento de sentenças, ou fragmentos de sentenças, delimitadas por ponto final, ou outros sinais de pontuação. O MIVES identifica e classifica sentenças versificadas completas, trechos iniciais ou finais. Uma sentença completa começa após um ponto final, ou início de parágrafo, e termina com um ponto final ou final de parágrafo. Um trecho inicial também começa em um ponto final e pode finalizar em qualquer sinal de pontuação, ponto final ou não. Um trecho final começa após qualquer sinal de pontuação e termina em um ponto final. A busca por trechos iniciais, ou finais, pode fornecer partes de uma sentença mas também sentenças completas delimitadas por pontos finais. Na etapa seguinte, ocorre a separação silábica ortográfica de cada palavra da sentença, e a escansão das sílabas poéticas considerando diversos fenômenos fonológicos, como a identificação de posicionamento das sílabas tônicas, fusões e separações inter-silábicas. Um exemplo de trecho inicial é ‘Sentou no fundo da igrarité virada’, obtido da sentença original ‘Sentou no fundo da igrarité virada, esperando.’ de *Macunaíma, o herói sem nenhum caráter*, de Mário de Andrade, 1928. O exemplo de uma sentença completa da mesma obra é ‘Acabou-se a história e morreu a vitória.’

Na etapa seguinte, ocorre a separação silábica ortográfica de cada palavra da sentença, e a escansão das sílabas poéticas, considerando vários fenômenos fonológicos, como a identificação de posicionamento das sílabas tônicas, fusões e separações inter-silábicas. ‘Escansão’ é o processo de contagem, identificação e classificação de sílabas poéticas. Em língua portuguesa, a

¹⁴O código fonte do PROPOE está disponível em <https://github.com/lasicuefs/propoe>

Sentença	Escansão	Métrica	Tônicas
Sentou no fundo da igarité virada,	Sen/t#o/u/ no/ f#un/do/ da i/ga/ri/t#é/ vi/r#a/da,	12	2 5 10 12
Acabou-se a história e morreu a vitória.	A/ca/b#ou-/se a his/t#ó/ri/a e/ mo/rr#e/u a/ vi/t#ó/ria.	12	3 5 9 12
Acabou-se a história e morreu a vitória.	A/ca/b#ou-/se a his/t#ó/ria e/ mo/rr#eu a/ vi/t#ó/ria.	10	3 5 8 10
Acabou-se a história e morreu a vitória.	A/ca/b#ou-/se a his/t#ó/ria e/ mo/rr#e/u a/ vi/t#ó/ria.	11	3 5 8 11
‘Ia pro céu viver com a marvada.	#I/a/ pr#o/ c#éu/ vi/v#er/ com/ a/ mar/v#a/da.	10	1 3 4 6 10
O herói não maliciou nada.	O he/r#ó/i/ n#ão/ ma/li/ci/#o/u/ n#a/da.	10	2 4 8 10
De certo era dinheiro enterrado.	De/ c#er/to/ #e/ra/ di/nh#e/i/ro/ en/te/rr#a/do	12	2 4 7 12
Passou um tempo de silêncio sagrado.	Pa/ss#o/u um/ t#em/po/ de/ si/l#ên/ci/o/ sa/gr#a/do	12	2 4 8 12

Tabela 3: Exemplos de sentenças extraídas pelo MIVES da obra *Macunaíma*

estrutura do verso é determinada pela relação entre as sílabas poéticas, e alternância de acentos até a última sílaba acentuada. Fenômenos fonológicos diversos podem unir ou separar as sílabas poéticas. No processo de identificação de sílabas, o MIVES considera fenômenos de sinérese, diérese, sinalefa, elisão e crase.¹⁵

Como o resultado da escansão não é unívoca, uma mesma sentença pode ser escandida pelo MIVES com diferentes medidas. O resultado é um conjunto de sentenças que atende aos limites mínimos e máximos de estrutura métrica estabelecidos pelo usuário, permitindo a exportação de sentenças identificadas na obra literária; cada sentença é associada à sua forma original e às suas possíveis versões escandidas com separação de sílabas poéticas e marcação de tônicas. A Tabela 3 exemplifica algumas sentenças e suas escansões obtidas pelo MIVES na obra *Macunaíma*.

3.2. Divisão e agrupamento das sentenças

O PROPOE recebe o resultado do MIVES e inicia o processamento separando as sentenças pela última sílaba da última palavra de cada sentença. Assim, todas as sentenças que terminam com a mesma sílaba final são agrupadas, facilitando a recuperação de sentenças por sílaba final para etapa de filtragem em seguida. Cada sentença tem associada a ela uma ou mais medidas de metro, resultado de sua escansão, marcação de tônicas e padrão rítmico. Assim, por exemplo, a partir de uma sílaba final ‘da’ e a lista de sentenças da Tabela 3 obtém-se a sub-lista de sentenças ‘Sentou no fundo da igarité virada’, ‘Ia pro céu viver com a marvada’ e ‘O herói não maliciou nada’ com suas escansões, métricas e esquema rítmico (posicionamento de

tônicas). Como a mesma sentença pode ser escandida diferentemente, ela pode aparecer muitas vezes na lista de sentenças para uma determinada sílaba. A sentença ‘Acabou-se a história e morreu a vitória’ pode, por exemplo, ser escandida com 10 sílabas (decassílabo), 11 sílabas (hendecassílabo), ou 12 (dodecassílabo), conforme Tabela 3, e cada versão da escansão é associada à sílaba final ‘te’.

3.3. Filtro das sentenças

Devido à grande quantidade de sentenças metrificadas que podem ser encontradas, o PROPOE realiza um processo de filtragem para reduzir o número de sentenças a serem manipuladas na montagem do poema. Este processo produz uma sub-lista de sentenças, com sentenças correspondentes à parametrização escolhida, minimizando o custo computacional. Para exemplificar, o número total de sentenças, ou trechos de sentenças, decassilábicas, identificadas em *Os Sertões* (Cunha, 1902) é 1.952. Existem, portanto, 14.518.416.572.416 possíveis arranjos de 4 sentenças. Deste modo, a montagem baseia-se em combinações de sentenças somente entre um número reduzido delas.

A filtragem de sentenças é feita por sílaba, por estrutura do poema e por métrica. Através do filtro silábico, duas ou quatro sílabas são escolhidas, produzindo um subconjunto de sentenças que possuem sílabas selecionadas como última sílaba. Este filtro reduz o número de sentenças a serem consideradas na produção de estrofes com rimas externas¹⁶ na fase de montagem. Um recurso utilizado para montagem do verso, que também afeta a filtragem de sentenças, é a escolha do número de estrofes do poema e do número de versos por estrofes. A partir desses parâmetros, são determinadas sentenças candidatas em quantidade que podem atender ao número total de versos na montagem do poema,

¹⁵Sinérese: união de duas vogais, formando uma unidade sonora. Diérese: separação de duas vogais em uma mesma sílaba. Sinalefa: transforma duas sílabas formadas por vogais em uma só sílaba. Elisão: união em uma sílaba métrica de diferentes vogais, em uma só emissão de voz. Crase: união em uma sílaba métrica de vogais idênticas. (Silva, 2014)

¹⁶A rima externa é a semelhança fonética entre duas palavras no final de dois versos. Já a rima interna, é a semelhança fonética entre duas palavras no meio dos versos (Goldstein, 2008).

evitando a repetição de uma mesma sentença no mesmo poema, por falta de sentenças distintas. Por exemplo, se a estrutura do poema tem duas estrofes de quatro versos considerando somente duas sílabas finais, então serão necessárias pelo menos quatro sentenças para cada sílaba.

O filtro de métrica permite selecionar um conjunto de sentenças a partir dos metros obtidos na escansão de sílabas poéticas. Os metros disponíveis no sistema PROPOE variam de uma a treze sílabas. Ao selecionar um metro específico, são filtradas somente as sentenças que tenham o metro associado, produzindo um poema com versos isométricos. Este filtro é opcional. Se não for escolhida um metro, todos os metros são considerados e é gerado um poema heterométrico.

3.4. Montagem dos versos

A montagem final do poema corresponde a organização de versos em estrofes a partir de uma lista de sentenças filtradas na etapa anterior. O problema computacional nesta fase está relacionado a escolha de versos entre as sentenças candidatas, já filtradas por sílaba, estrutura do poema e métrica conforme etapa anterior, e posicionamento dos versos em estrofes. O PROPOE trata a seleção e a ordenação de sentenças como um problema de otimização combinatória de produção de um arranjo de k sentenças entre n possíveis, de modo a maximizar critérios rítmicos do poema. Uma estratégia possível para esta otimização combinatorial poderia testar todas as combinações para obter o melhor padrão rítmico no poema, ou seja, aplicando uma estratégia de otimização global ao poema. Porém, teria um alto custo computacional avaliar todas as combinações, mesmo após a filtragem, dado o grande número de sentenças possíveis.

A abordagem do PROPOE aplica uma técnica ‘dividir-e-conquistar’ (*divide-and-conquer*) e ‘gulosa’ (*greedy*), para reduzir o custo computacional (Cormen et al., 2009). Através de divisão e conquista, é feita uma fragmentação do problema de geração de um poema em partes. Em seguida, as soluções obtidas para as estrofes são combinadas para obter uma solução geral. Já as estrofes são geradas através de uma abordagem gulosa, que busca otimizar localmente os critérios rítmicos, comparando somente pares de versos, para efetuar a seleção de um verso por vez em cada posição.

A montagem do poema é apresentada na Figura 2. Ela tem início através de um conjunto de sentenças já filtradas em uma etapa anterior, e baseia-se nos parâmetros: escolha de sílabas fi-

nais, número de estrofes e de versos por estrofe, tipos de rima externa, critério de comparação e similaridade. O poema resulta da justaposição sequencial das estrofes montadas.

3.4.1. Montagem da Estrofe

Cada estrofe é montada separadamente e de acordo com os parâmetros fornecidos. A quantidade de sílabas finais e de estrofes determinam quais sentenças candidatas são exploradas em cada estrofe. Se são escolhidas duas sílabas finais, os dois conjuntos de sentenças associadas a essas sílabas são utilizadas em todas as estrofes. Se são escolhidas quatro sílabas, as estrofes ímpares utilizam as sentenças das 1^a e 2^a sílabas e as estrofes pares utilizam as sentenças das 3^a e 4^a sílabas.

O primeiro verso da estrofe é escolhido aleatoriamente entre as sentenças candidatas terminadas com uma das sílabas finais escolhidas pelo usuário. Este primeiro verso torna-se uma referência que regula a seleção dos demais versos da estrofe, através de sucessivas comparações rítmicas. O posicionamento dos versos seguintes depende da escolha do tipo de rima externa (homofonia externa) para as estrofes. No português, a rima externa ocorre através da repetição de sons no final dos versos, que é o parâmetro estrutural mais facilmente identificado. Nesta etapa, consideramos que a rima entre versos ocorre quando as sentenças candidatas possuem a mesma última sílaba poética.¹⁷

Para estrofes em quadras (quatro versos), o PROPOE pode gerar poemas de rima cruzada, com versos alternados ABAB, onde A e B correspondem a uma sentença com uma sílaba final A ou B; rima emparelhada, mantendo uma sequência de cada dois versos, AABB; rima interpolada, ao rimar o primeiro verso com o quarto, e o segundo com o terceiro, ABBA; rima mista, formado sem imposição de ordenação pré-definida dos versos na estrofe. Na rima mista, é sorteado, a cada verso, se ele é do tipo A ou B, expandindo as possibilidades de combinações entre as rimas, podendo gerar AAAA, AAAB, AABA, ABAA, BAAA, BBBB, BBBA, BBAB, ou qualquer combinação de rimas anteriormente descritas. Em poemas de rima cruzada, emparelhada ou interpolada, um fator que pode alterar o resultado final é o número de versos por estrofe. Embora o PROPOE gere poemas com variado número de versos por estrofe, a rima baseia-se em uma es-

¹⁷Um caso específico de sentenças com a mesma sílaba final seriam sentenças com a mesma palavra final, e esta possibilidade não é excluída durante a montagem.

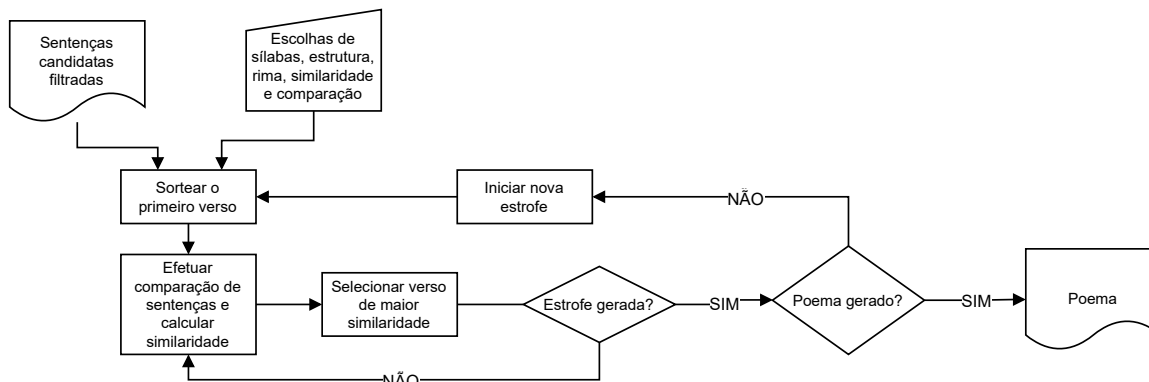


Figura 2: Etapas da montagem do poema.

trutura de quartetos. Se a estrutura é menor, o padrão é interrompido antes de ser completado; por exemplo, um terceto de rima cruzada ABAB é interrompido em ABA. Da mesma forma, se a estrutura é maior, o padrão é repetido total ou parcialmente; assim, um sexteto de rima emparelhada AABB é AABBA.

Se o primeiro verso de cada estrofe é selecionado aleatoriamente, os demais são escolhidos por similaridade rítmica com o verso de referência. O verso de referência pode ser o primeiro verso da estrofe ou o verso imediatamente anterior, conforme escolha do usuário para este parâmetro da montagem. Para evitar a repetição de versos, cada sentença posicionada no poema é marcada como já utilizada na lista de sentenças candidatas para não ser utilizada em novas seleções. Após comparação das sentenças candidatas com o verso de referência, o verso selecionado é aquele que possui maior similaridade rítmica com ele. Como parâmetro do sistema, esse critério pode ser de similaridade por padrão rítmico ou de acentuação da última tônica. O padrão rítmico corresponde à distribuição posicional das sílabas fortes (tônicas) na sentença; a regra de similaridade da acentuação corresponde à posição da última sílaba tônica nas palavras que rimam. É efetuado um cálculo de similaridade de acordo com o critério escolhido para determinar a maior similaridade entre as sentenças candidatas. Em caso de empate, o outro critério é utilizado.

Para obter uma similaridade de padrão rítmico, o PROPOE utiliza o ‘coeficiente de Jaccard’ para comparar os conjuntos U e V de posições de tônicas de cada verso:

$$Jaccard(U, V) = \frac{|U \cap V|}{|U \cup V|} \quad (1)$$

O coeficiente de Jaccard mede a similaridade entre dois conjuntos finitos e é definido como a razão entre a cardinalidade da interseção dos con-

juntos e a cardinalidade da união dos conjuntos. Se considerarmos o esquema rítmico de um verso como um conjunto de números correspondentes, a posição das sílabas tônicas de cada verso, é possível calcular o coeficiente de Jaccard entre esquemas rítmicos e assim obter um valor de similaridade entre eles. Se o coeficiente tiver o valor 1.0, os esquemas são iguais; se o coeficiente é 0.0, não há qualquer posição tônica similar.

Como exemplo, as sentenças ‘De certo era dinheiro enterrado’ e ‘Passou um tempo de silêncio sagrado’, da Tabela 3, tem tônicas nas posições $\{2, 4, 7, 12\}$ e $\{2, 4, 8, 12\}$, respectivamente. A interseção das posições das tônicas então seria $\{2, 4, 12\}$ enquanto a união seria $\{2, 4, 7, 8, 12\}$, e assim a similaridade de padrão rítmico seria $3/5 = 0.6$.

Quanto à similaridade de acentuação da última tônica, o PROPOE determina para cada verso qual a posição da sílaba tônica na última palavra. A acentuação pode ser aguda (oxítone), grave (paroxítone) ou esdrúxula (proparoxítone). Se duas sentenças têm a mesma acentuação, sua casa similaridade é 1.0, caso contrário a similaridade é 0.0. Considerando como exemplo as mesmas sentenças anteriores, ambas têm acentuação grave (paroxítone), logo a similaridade é de 1.0.

O processo de montagem das estrofes é regulado por um ‘procedimento guloso’, uma vez que as comparações de similaridade são feitas em pares, de maneira iterativa, e não entre todas as sentenças da estrofe, ou do poema, que produziria um custo computacional muito maior. O poema final é resultante portanto da associação de estrofes construídas por esta abordagem de otimização local. Uma avaliação global das características rítmicas do poema é o último passo do processo de sua geração.

3.5. Avaliação do poema

A avaliação automatizada de poemas é um problema abordado em muitos trabalhos, como indica [Gonçalo Oliveira \(2017b\)](#), que concentra-se em dimensões objetivas de avaliação. A avaliação produzida pelo PROPOE é realizada pelo próprio sistema segundo padrões rítmicos. Para isso, é feita uma avaliação com base em critérios que calculam as pontuações, normalizadas no intervalo $[0; 1]$, somando-as para obter uma avaliação final. Como são avaliados quatro critérios no poema, a pontuação total pode variar de 0.0 até 4.0 pontos. Os critérios estabelecidos para avaliação do poema estão relacionados à similaridade do esquema rítmico, correspondência das sílabas nas posições de tônicas, acentuação de última tônica, e similaridade entre as palavras que rimam. Cada um destes critérios contribuem para a avaliação do poema.

Sobre o critério de similaridade do esquema rítmico, é analisado o padrão rítmico geral do poema que contribui para sua avaliação porque produz um padrão de variação de intensidade sonora, de acordo com a posição das sílabas átonas e tônicas nos versos. É feita a comparação e cálculo de similaridade de Jaccard do esquema rítmico de cada verso com o verso de referência, de maneira análoga ao cálculo de similaridade de padrão rítmico efetuado na montagem. Em seguida, é feita a soma dos coeficientes encontrados para todos os versos de todas as estrofes do poema. Para normalizar este valor, esta soma é dividida pelo número de comparações realizadas. Por exemplo, se um poema tiver duas estrofes com quatro versos cada, serão realizadas seis comparações para avaliação de similaridade, uma vez que cada verso é avaliado somente uma vez, comparando com o verso de referência e o primeiro verso de cada estrofe não tem verso de referência.

O critério relacionado às sílabas tônicas idênticas avalia outro efeito sonoro decorrente de sílabas idênticas que ocupam a mesma posição tônica nos versos. Duas sílabas tônicas são consideradas idênticas se possuem as mesmas letras e estão na mesma posição. Feita a identificação das sílabas tônicas idênticas, o poema recebe um ponto para cada sílaba igual identificada. O resultado da soma destas pontuações é dividida pelo número de comparações realizadas para normalizar a pontuação. Por exemplo, ao comparar os versos ‘A palha o fosfre e o goiano’ (A/p#a/lha/ o/ f#os/fre/ e o/ go/i/#a/no) e ‘Faz com que os ventos do oceano’ (F#az/ com/ que/ os/ v#en/tos/ do/ o/ce/#a/no), há duas tônicas em mesma posição, {5, 10}, e a sílaba tônica

na posição 10, /#a/, é coincidente então a pontuação desta comparação será $1/2=0.5$.

Outro critério de avaliação está relacionado à acentuação das últimas palavras nos versos. Há uma pequena diferença de referência, neste caso, em relação ao critério de acentuação para montagem. Para fins de avaliação, a comparação da acentuação é feita com base nas palavras que rimam, ou seja, cada verso é comparado com o verso anterior de mesma rima. Se duas sentenças têm a mesma acentuação, sua casa similaridade é 1.0, caso contrário a similaridade é 0.0. O número de coincidências de acentuação existentes em um poema é dividido pelo total de comparações realizadas para obter a pontuação.

O último critério de avaliação corresponde à similaridade de letras das palavras que rimam ao final de cada verso. Assim, duas palavras finais de versos que rimam na estrofe são comparadas, letra a letra, em sentido inverso, e cada correspondência de letra é avaliada. Esta contagem é dividida pelo tamanho da menor das palavras comparadas. Por exemplo, ao comparar a palavra final ‘sagrado’ de um verso com a palavra final ‘enterrado’ de outro verso, há três letras finais iguais e a menor palavra tem sete letras, então a pontuação desta comparação será $3/7=0.43$. Tal procedimento é aplicado para cada par de versos que rimam, somando-se os valores obtidos a cada comparação. Em seguida, o valor total da soma é dividido pelo número de versos comparados.

As pontuações obtidas por cada um dos quatro critérios de avaliação, ao fim, são somadas e este valor define a pontuação do poema.

4. Resultados

O PROPOE, como descrito, gera poemas isométricos e heterométricos (deca e dodecasilábicos), baseados em estruturas rítmicas regulares, a partir de sentenças metrificadas extraídas de obras da prosa em língua portuguesa, pelo MIVES. A geração baseia-se na aplicação de filtros e na montagem de sentenças a partir de padrões rítmicos estabelecidos. O poema é avaliado pelo próprio sistema, através de pontuações relacionadas ao padrão rítmico e fonético (rima), e acentuação.

Selecionamos quatro obras da literatura brasileira: *Os Sertões*, de Euclides da Cunha, 1902; *Macunaíma, o herói sem nenhum caráter*, de Mário de Andrade, 1988; *Serafim Ponte Grande*, de Oswald de Andrade, 1933; *Grande Sertão: Veredas*, de Guimarães Rosa, 1956. Elas foram processadas pelo MIVES, e os resultados são apresentados na Tabela 4, com um resumo descritivo

Obra	Total de Sentenças	Tipo de Sentença	Sentenças Métricas (6 a 13)	Decassílabos	Dodecassílabos
<i>Os Sertões</i>	8.361	Completas	3.182	409	414
		Trechos Iniciais	5.391	912	1023
		Trechos Finais	5.838	730	1284
<i>Serafim Ponte Grande</i>	1803	Completas	828	115	111
		Trechos Iniciais	926	198	76
		Trechos Finais	1172	210	205
<i>Macunaíma</i>	2897	Completas	1047	154	149
		Trechos Iniciais	1596	269	250
		Trechos Finais	1714	225	355
<i>Grande Sertão: Veredas</i>	14725	Completas	6401	954	912
		Trechos Iniciais	12481	2006	2264
		Trechos Finais	12122	1433	2544

Tabela 4: Quantidade de sentenças métricas nas obras selecionadas.

Variações	Quantidade de Poemas	Pontuação				
		Esquema Rítmico	Sílabas Tônicas Iguais	Acentuação	Letras Finais Iguais	Geral
Obra Literária						
<i>Os Sertões</i>	25	0.65±0.09	0.03±0.05	1.00±0.02	0.57±0.09	2.24±0.19
<i>Serafim Ponte Grande</i>	25	0.50±0.11	0.01±0.02	0.89±0.20	0.45±0.16	1.86±0.30
<i>Macunaíma</i>	25	0.58±0.11	0.03±0.04	1.00±0.00	0.48±0.14	2.10±0.22
<i>Grande Sertão Veredas</i>	25	0.69±0.12	0.06±0.07	1.00±0.00	0.67±0.14	2.41±0.18
Número de Estrofes						
1	32	0.57±0.14	0.02±0.04	0.92±0.18	0.52±0.20	2.03±0.39
2	57	0.63±0.12	0.03±0.04	0.99±0.04	0.56±0.13	2.21±0.23
3	11	0.59±0.12	0.05±0.09	1.00±0.00	0.55±0.15	2.19±0.22

Tabela 5: Avaliação de poemas gerados pelo PROPOE com pontuação média e desvio padrão.

das sentenças metrificadas entre hexassílabos e bárbaros (métricas de seis a treze)). É importante salientar que uma mesma sentença pode ser escandida de diferentes formas, podendo aparecer na lista com diferentes medidas.

As sentenças metrificadas identificadas pelo MIVES, nestas obras, foram enviadas ao PROPOE para geração de poemas, através de variações métricas de versificação, número de estrofes, número de versos por estrofe, tipos de rima, uso de sentenças completas ou trechos iniciais e finais. Para avaliar os resultados da geração de poemas pelo PROPOE, foram produzidos 100 poemas, 25 poemas para cada obra com variações estruturais. Os resultados de pontuação de avaliação estão resumidos na Tabela 5.

Uma análise geral das pontuações dos 100 poemas gerados indica que a pontuação geral dos poemas está, em maior parte, entre 1.8 e 2.6. Há, no entanto, uma diferenciação da pontuação

geral a depender da obra ou do tamanho do poema. Obras com maior número de sentenças candidatas, como *Os Sertões* e *Grande Sertão: Veredas*, possuem pontuação geral média maior do que obras mais curtas, como *Macunaíma* e *Serafim Ponte Grande*. Pode ser esperada uma relação entre quantidade de sentenças candidatas e a pontuação geral, uma vez que quanto maior a oferta de possibilidades para combinações, maior a chance de obter uma sentença que maximize critérios rítmicos. Quanto ao tamanho do poema gerado, em número de estrofes, a pontuação geral média de poemas curtos (1 estrofe) tende a ser inferior aos poemas mais longos (2 ou 3 estrofes). Como as obras mais curtas oferecem menor número de sentenças candidatas, é mais difícil obter poemas longos pela falta de versos suficientes. Então, grande parte dos poemas mais longos resultam de obras com mais sentenças candidatas porque permitem mais possibilidades de otimização.

A análise dos critérios individuais de avaliação evidenciam que há uma diferenciação da tendência de pontuação conforme o critério. Há um critério mais fácil de ser atendido, que é a pontuação por semelhança de acentuação, que tende a pontuações mais altas, possivelmente devido a pouca variação (três possibilidades) de posicionamento da tônica na última palavra. Já a pontuação por sílabas tônicas iguais tem valores consistentemente baixos, indicando que este é um critério difícil de ser satisfeito, por não ser tratado explicitamente durante a montagem. Por outro lado, as pontuações por similaridade de esquema rítmico e por letras iguais na última palavra tendem a pontuações intermediárias, uma vez que são considerados, ao menos parcialmente, na montagem. Para exemplificar os resultados obtidos na geração e pontuação de poemas individuais pelo sistema, com diferentes parametrizações, alguns foram selecionados para análise do processo de montagem e poemas.

Os três primeiros poemas são mais curtos, de uma única estrofe de 4 versos (quadra), com diversos tipos de sentenças (completas ou extrações parciais), métricas, parâmetros de filtragem e montagem. O Poema 1 é uma quadra gerada a partir de sentenças completas de *Macunaíma*, filtradas para obter somente versos dodecassilábicos, com sílabas finais ‘do’ e ‘da’ (Tabela 6). O PROPOE recebeu 151 sentenças completas, processadas do *Macunaíma*, sendo 88 terminadas com ‘do’ e 44 com ‘da’. A filtragem dodecassilábica reduziu este número para 14 sentenças terminadas com ‘do’ e 8 com ‘da’. A montagem foi feita para uma estrutura de uma estrofe de quatro versos, com rima emparelhada (AABB), seguindo um critério de similaridade por padrão rítmico (posição das tônicas), e comparação com o primeiro verso da estrofe para seleção de cada verso subsequente.

Para exemplificar o funcionamento das etapas de montagem, descrevemos os passos seguidos até obter este poema. O primeiro verso da estrofe do

Poema 1 foi sorteado entre as sentenças candidatas. Uma vez que a primeira sílaba sorteada foi ‘do’, foi sorteado o verso ‘De certo era dinheiro enterrado’ para esta posição. Para selecionar o verso seguinte, outras 13 sentenças disponíveis terminadas com ‘do’ foram comparadas com o primeiro verso, através de cálculo de similaridade do padrão rítmico. O primeiro verso possui um padrão de posição de tônica {2, 4, 7, 12}. Entre as opções candidatas de maior similaridade, encontra-se o ‘Passou um tempo de silêncio sagrado.’ (Pa/ss#o/u um/ t#em/po/de/ si/l#ên/ci/o/ sa/gr#a/do) com tônicas nas sílabas {2, 4, 8, 12} e o ‘Porém Oibê não fez caso e veio vindo.’ (Po/r#ém/ Oi/b#ê/ n#ão/ f#ez/ c#a/so e/ v#e/i/o/ v#in/do.) com tônicas em {2, 4, 5, 6, 7, 9, 12}. Como o primeiro verso candidato tem similaridade de 0.6, com relação ao verso de referência, e o segundo verso candidato tem similaridade de 0.57, a primeira opção foi selecionada. A terceira posição da estrofe é preenchida pela seleção entre as oito sentenças terminadas com ‘da’, considerando a rima emparelhada, e avaliando similaridade de esquema rítmico com o primeiro verso da estrofe. O verso candidato ‘Pulou nesse e abriu na galopada’ (Pu/l#o/u/ n#e/sse/ e a/br#i/u/ na/ga/lo/p#a/da) tem padrão {2, 4, 7, 12} e similaridade máxima de 1.0, com todas as tônicas em posições idênticas ao do primeiro verso, e selecionado para a terceira posição. Em seguida, outro verso terminado em ‘da’, ‘A companheira não trabalhara nada’ (A/ com/pa/nh#e/i/ra/ n#ão/ tra/ba/lh#a/ra/ n#a/da) com tônicas nas posições {4, 7, 10, 12} obtendo similaridade de 0.6, maior valor entre as sentenças restantes.

Após a montagem, o Poema 1 foi avaliado pelo sistema obtendo score de 2.39, de um máximo de 4.0, na soma dos quatro critérios. A pontuação de similaridade do esquema rítmico foi 0.73, obtida pela média das similaridades das posições das tônicas entre os versos comparados com o verso de referência (neste caso, o pri-

De certo era dinheiro enterrado
 Passou um tempo de silêncio sagrado
 Pulou nesse e abriu na galopada
 A companheira não trabalhara nada

escansão	métrica	tônicas
De/ c#er/to/ #e/ ra/ di/nh#e/i/ro/ en/te/rr#a/do	Dodecassílabo	2 4 7 12
Pa/ss#o/u um/ t#em/po/ de/ si/l#ên/ci/o/ sa/gr#a/do	Dodecassílabo	2 4 8 12
Pu/l#o/u/ n#e/sse/ e a/br#i/u/ na/ ga/lo/p#a/da	Dodecassílabo	2 4 7 12
A/com/pa/nh#e/i/ra/ n#ão/ tra/ba /lh#a/ra/ n#a/da	Dodecassílabo	4 7 10 12

Tabela 6: Poema 1 gerado pelo PROPOE a partir de *Macunaíma*.

De noite continuou chorando
Agora o herói está fatigado
A palha o fosfre e o goiano
Faz com que os ventos do oceano

escansão	métrica	tônicas
De/ n#oi/te/ con/ti/nu/#o/u/ cho/r#an/do	Decassílabo	2 7 10
A/g#o/ra/ o he/r#ó/i es/t#á/ fa/ti/g#a/do	Decassílabo	2 5 7 10
A/ p#a/lha/ o/ f#os/fre/ e o/ go/i/#a/no	Decassílabo	2 5 10
F#az/ com/ que/ os/ v#en/tos/ do/ o/ce/#a/no	Decassílabo	1 5 10

Tabela 7: Poema 2 gerado pelo PROPOE a partir de *Macunaíma*.

meiro) – 0.6, 1.0 e 0.6. A pontuação por sílabas tônicas idênticas foi 0.0, o que reforça a dificuldade de pontuar neste critério, uma vez que não há sílabas idênticas na mesma posição (tônica) em comparação com o verso de referência. A pontuação por acentuação das últimas palavras obteve pontuação máxima de 1.0, resultado da comparação da acentuação das palavras que rimam, uma vez que todos os versos têm acentuação grave (tônica na penúltima sílaba). A pontuação por similaridade de letras das últimas palavras foi 0.66, acima da média dos poemas advindos da obra *Macunaíma*, pois ‘sagrado’ e ‘enterrado’ têm quatro letras idênticas (em ordem inversa); ‘nada’ e ‘galopada’ possuem três letras idênticas.

O Poema 2 (Tabela 7) também obteve sentenças completas de *Macunaíma* para gerar uma quadra. Mas, diferente do Poema 1, foram filtradas sentenças decassilábicas, sílabas finais ‘do’ e ‘no’, e alterado o verso de referência de similaridade para o verso imediatamente anterior. A filtragem decassilábica forneceu 15 sentenças que terminaram com ‘do’, e 4 sentenças com ‘no’. A montagem foi configurada para uma estrutura de 1 estrofe de quatro versos, rima emparelhada (AABB), e critério de similaridade por padrão rítmico (posição das tônicas) para comparação com o verso imediatamente anterior, regulando a seleção de cada verso subsequente.

A avaliação do Poema 2, após a montagem, indicou um escore total de 2.15. A pontuação de similaridade do esquema rítmico foi 0.67, considerando a média das pontuações 0.75, 0.75 e 0.5, nas comparações de cada sentença com a anterior. Havia muitas sentenças para a primeira sílaba final, levando a pontuações mais altas, porém poucas para a segunda sílaba final, limitando a otimização deste critério. A pontuação por sílabas tônicas idênticas foi 0.11, um valor maior que o obtido no poema anterior, já que, entre as nove comparações de letras de sílabas tônicas, foi identificada uma sílaba tônica idêntica, ‘a’ entre o terceiro e o quarto versos

na décima posição. A pontuação por acentuação das últimas palavras obteve pontuação máxima de 1.0 já que todos os versos têm acentuação grave. E a pontuação por similaridade de letras das últimas palavras obteve pontuação de 0.38, pois ‘chorando’ e ‘fatigado’ possuem três letras idênticas, como ‘goiano’ e ‘oceano’.

Serafim Ponte Grande, de Oswald de Andrade, forneceu sentenças completas para geração do Poema 3 (Tabela 8). Foram filtradas sentenças dodecassilábicas, terminadas com ‘do’ e ‘te’, obtendo sete candidatas para cada sílaba. A montagem novamente foi configurada para uma quadra de rima emparelhada, seguindo um critério de similaridade por padrão rítmico (posição das tônicas) em comparação com o verso imediatamente anterior. A montagem selecionou três versos alexandrinos (tônicas em 6 e 12). Na avaliação final, o poema obteve escore de 2.11. A pontuação de similaridade do esquema rítmico foi 0.59, considerando a média das pontuações 0.67, 0.67 e 0.43, nas comparações de similaridade de posicionamento das tônicas de cada sentença com a anterior. Não foi obtida qualquer sílaba tônica idêntica na comparação de sentenças, com pontuação zero nesse critério. Obteve-se pontuação máxima de acentuação da última palavra, já que todos os versos terminam com acentuação grave. Para a similaridade de letras das palavras que rimam, a pontuação foi 0.52, já que ‘frente’ e ‘silente’ possuem quatro letras finais idênticas (e a menor palavra tem seis letras), e ‘enrugado’ e ‘atentado’ possuem três letras finais iguais (ambas possuem oito letras).

O Poema 4 (Tabela 9) baseou-se em *Os Sertões*, de Euclides da Cunha. Foram consideradas sentenças finais. A filtragem selecionou sentenças decassilábicas, terminadas com as sílabas finais ‘dos’ e ‘tes’, e obteve 24 e 22 candidatas para cada sílaba, respectivamente, uma quantidade maior de possibilidades do que aquelas exibidas nos cenários anteriores. Foi montada uma quadra de rima emparelhada aplicando o

Nosso herói tende ao anarquismo enrugado
 Contra ele preparo um imenso atentado
 Tudo conspira nesta cidade silente
 O meu relógio anda sempre para a frente

escansão	métrica	tônicas
N#o/sso/ he/r#ó/i/ t#en/de ao a/nar/qu#is/mo en/ru/g#a/do	Dodecassílabo	1 4 6 9 12
C#on/tra/ #e/le/ pre/p#a/ro um/ i/m#en/so a/ten/t#a/d	Dodecassílabo	1 3 6 9 12
T#u/do/ cons/p#i/ra/ n#es/ta/ ci/d#a/de/ si/l#en/te	Dodecassílabo	1 4 6 9 12
O/ m#eu/ re/l#ó/gi/o/ #an/da/ s#em/pre/ para a/ fr#en/te	Dodecassílabo	2 4 7 9 12

Tabela 8: Poema 3 gerado pelo PROPOE a partir de *Serafim Ponte Grande*.

critério de similaridade por padrão rítmico, em comparação com o verso imediatamente anterior. Em razão da grande quantidade de sentenças candidatas, foi obtida uma maior variedade de posicionamento de tônicas. A busca por similaridade obteve todos os versos exatamente com o mesmo esquema rítmico {2, 7, 10}. Assim, na avaliação final, de 2.52, a pontuação de similaridade do esquema rítmico contribuiu com o valor 1.0, evidenciando que a maior quantidade de sentenças candidatas pode permitir melhor otimização deste critério. Entretanto, a ausência de sílabas tônicas idênticas na comparação de sentenças recebeu uma pontuação zero nesse critério; ou seja, mesmo com muitas sentenças, este critério permanece difícil de pontuar. A pontuação foi máxima para acentuação das últimas palavras, todas graves. A similaridade de letras das palavras que rimam produziu uma pontuação de 0.52, já que ‘verdejantes’ e ‘retumbantes’ possuem cinco letras finais idênticas, e ‘unidos’ e ‘isolados’ possuem três.

O Poema 5 (Tabela 10) foi obtido de *Os Sertões* utilizando sentenças completas para construir dois tercetos de dodecassílabos de estrutura rítmica ABB. As sílabas finais escolhidas foram ‘va’ e ‘so’, em 5 e 7 sentenças candidatas dodecassilábicas, respectivamente. Mesmo em uma obra com grande quantidade total de sentenças candidatas, há sílabas com quanti-

dade limitada de possibilidades. A montagem foi feita por comparação de similaridade de esquema rítmico, considerando a última palavra de cada verso, utilizando como referência o verso anterior. Como não havia sentenças métricas com as mesmas posições de tônicas, foram selecionadas as opções mais similares para geração do poema. A avaliação final do poema foi 2.15, com 1.0 ponto pela acentuação, já que são graves todas as palavras finais. Como não foram encontradas sílabas tônicas idênticas, não foi pontuado este critério. A avaliação fonética examinou apenas as sentenças terminadas com ‘so’, já que nesta estrutura de três versos por estrofe não há outra sentença com ‘va’. Assim, a pontuação foi 0.56 considerando três letras finais idênticas entre ‘nervoso’ e ‘desvalioso’, e duas letras finais idênticas entre ‘acaso’ e ‘lastimoso’. A avaliação de esquema rítmico obteve pontuação intermediária, de 0.58, valor esperado em consequência da limitação de sentenças candidatas.

A partir de sentenças completas de *Grande Sertão: Veredas*, foi gerado o Poema 6 (Tabela 11) com dois quartetos de decassílabos, rimas cruzadas, e sílabas finais ‘dos’ (17 sentenças candidatas) e ‘go’ (16 sentenças candidatas). A montagem foi feita por comparação de similaridade de esquema rítmico, utilizando como referência o verso anterior. O poema possui versos de padrões rítmicos clássicos — dois

ao cabo, completamente isolados
 perdiam-se nos casebres unidos
 as linhas de ordens do dia retumbantes
 ilhados, os igapós verdejantes

escansão	métrica	tônicas
ao/ c#a/bo./ com/ple/ta/m#en/te i/so/l#a/dos	Decassílabo	2 7 10
per/d#i/am-/se/ nos/ ca/s#e/bres/ u/n#i/dos	Decassílabo	2 7 10
as/ l#i/nhas/ de or/dens/ do/ d#ia/ re/tum/b#an/tes	Decassílabo	2 7 10
i/lh#a/dos,/ os/ i/ga/p#ós/ ver/de/j#an/tes	Decassílabo	2 7 10

Tabela 9: Poema 4 gerado pelo PROPOE a partir de *Os Sertões*.

Fez-se, porém, uma última tentativa
Determinou-a incidente desvalioso
Logo depois vibrava um abalo nervoso

Servidão inconsciente; vida primitiva
Uauá patenteava quadro lastimoso
Tomemos um fato, entre muitos, ao acaso

escansão	métrica	tônicas
F#ez/se,/ po/r#ém,/ #u/ma/ #úl/ti/ma/ ten/ta/t#i/va	Dodecassílabo	1 4 5 7 12
De/ter/mi/n#o/ua in/ci/d#en/te/ des/va/li/#o/so	Dodecassílabo	4 7 12
L#o/go/ de/p#ois/ vi/br#a/va um/ a/b#a/lo/ ner/v#o/so	Dodecassílabo	1 4 6 9 12
Ser/vi/d#ão in/cons/ci/#en/te;/ v#i/da/ pri/mi/t#i/va	Dodecassílabo	3 6 8 12
U/au#á/ pa/ten/te/#a/va/ qu#a/dro/ las/ti/m#o/so	Dodecassílabo	2 6 8 12
To/m#e/mos/ um/ f#a/to, #en/tre/ m#u/i/tos,/ ao a/c#a/so	Dodecassílabo	2 5 6 8 12

Tabela 10: Poema 5 gerado pelo PROPOE a partir de *Os Sertões*.

decassílabos heróicos (tônicas em 6 e 10), um martelo (tônicas em 3, 6 e 10), e versos de gaita galega (decassílabos com tônicas em 4, 7 e 10). Ele obteve pontuação final de 2.55. A acentuação teve 1.0, já que todas palavras finais são graves, confirmando a facilidade de pontuação máxima neste critério. A avaliação fonética teve uma pontuação 0.73, já que ‘digo’ e ‘contigo’ possuem 3 letras finais idênticas, enquanto os demais pares de palavras têm somente a sílaba final com rima idêntica. A avaliação de sílabas tônicas idênticas teve pontuação de 0.08, em ‘Eu’ no segundo e terceiro versos e ‘re’ no primeiro e segundo versos da segunda estrofe, indicando que é possível alguma

melhora nesse difícil critério, mas a pontuação tende a ser baixa devido a quantidade de tônicas diferentes. A pontuação de esquema rítmico obteve valor final de 0.73 com muitas intersecções de posições de tônicas em comparação com o verso anterior, em uma situação de maior quantidade de sentenças disponíveis.

Grande Sertão: *Veredas* também forneceu sentenças para o Poema 7 (Tabela 12), completas e parciais finais, em uma estrutura de dois quartetos, em versos dodecassílabos e ‘rima mista’, resultando numa escolha aleatória para sílaba final de cada verso, em um padrão ABAA e AAAA. As sílabas finais escolhidas foram ‘te’ e ‘to’, com

E uns outros, muito estampidos
Eu olhava para a beira do rego
Eu sabia do respirar de todos
Minha cara estava pegando fogo

Depois, tirei a dureza dos dedos
Calado, considerei comigo
Passei quase para a frente de todos
Ao dôido, doideiras digo

escansão	métrica	tônicas
E/ uns/ #o/u/tros,/ m#u/i/to es/tam/p#i/dos	Decassílabo	3 6 10
Eu/ o/lh#a/va/ para a/ b#e/i/ra/ do/ r#e/go	Decassílabo	1 3 6 10
Eu/ sa/b#i/a/ do/ res/pi/r#ar/ de/ t#o/dos	Decassílabo	1 3 8 10
M#i/nha/ c#a/ra es/t#a/va/ pe/g#an/do/ f#o/go	Decassílabo	1 3 5 8 10
De/p#ois,/ ti/r#e/i a/ du/r#e/za/ dos/ d#e/dos	Decassílabo	2 4 7 10
Ca/l#a/do,/ con/si/de/r#e/i/ co/m#i/go	Decassílabo	2 7 10
Pa/ss#e/i/ qu#a/se/ para a/ fr#en/te/ de/ t#o/dos	Decassílabo	2 4 7 10
Ao/ d#ô/i/do,/ do/i/d#e/i/ras/ d#i/go	Decassílabo	2 7 10

Tabela 11: Poema 6 gerado pelo PROPOE a partir de *Grande Sertão: Veredas*.

e o risco extenso dágua, de parte a parte
o senhor se entende? Deixe avante; conto
A razão dele era do estilo acinte
reconheço, ele tem tido uma sorte

sem prazo, isto é, o quase calados, somente
naquela sanha, é ver cachorrada caçante
por medo ter? Eu não campeava a morte
que carreguei o rifle, escorei, repetente

escansão	métrica	tônicas
e/ o/ r#is/co ex/t#en/so/ d#á/gua./ de/ p#ar/te a/ p#ar/te	Dodecassílabo	3 5 7 10 12
o/ se/nh#or/ se en/t#en/de?/ D#e/i/xe a/v#an/te;/ c#on/to	Dodecassílabo	3 5 7 10 12
A/ ra/z#ã/o/ d#e/le/ #e/ra/ do es/t#i/lo a/c#in/te	Dodecassílabo	3 5 7 10 12
re/co/nh#e/ço./ #e/le/ t#em/ t#i/do/ #u/ma/ s#or/te	Dodecassílabo	3 5 7 8 10 12
sem/ pr#a/zo./ #is/to é, o/ qu#a/se/ ca/l#a/dos./ so/m#en/te	Dodecassílabo	2 4 6 9 12
na/qu#e/la/ s#a/nha, é/ v#er/ ca/cho/rr#a/da/ ca/ç#an/te	Dodecassílabo	2 4 6 9 12
por/ m#e/do/ t#er?/ #Eu/ n#ão/ cam/pe/#a/va/ a/ m#or/te	Dodecassílabo	2 4 5 6 9 12
que/ ca/rre/gu#e/i o/ r#i/fle, es/co/r#ei./ re/pe/t#en/te	Dodecassílabo	4 6 9 12

Tabela 12: Poema 7 gerado pelo PROPOE a partir de *Grande Sertão: Veredas*.

um grande número de sentenças candidatas, respectivamente, 75 e 68. Isso permitiu gerar uma grande variedade de sentenças candidatas distintas para o poema. A sílaba ‘te’ forneceu seis sentenças, um quantitativo maior do que em outros exemplos, devido ao padrão de rima, uma exigência muito maior que a observada nos poemas anteriores. Nota-se que ainda assim foi possível obter uma segunda estrofe de versos alexandrinos. Sua pontuação final foi 2.52. O esquema rítmico contribuiu com 0.91 pontos, devido a repetição de posicionamento de tônicas, mas não obteve pontuação máxima como no Poema 4, apesar do número muito maior de sentenças disponíveis, o que indica que mesmo com grande quantidade de sentenças há desafios na otimização de critérios rítmicos. A avaliação fonética teve pontuação de 0.58 pontos, já que as palavras ‘somente’ e ‘caçante’ tiveram apenas uma letra idêntica, além da sílaba final. A acentuação teve 1.0 uma vez que todas palavras finais são graves; a avaliação de sílabas tônicas iguais teve pontuação de 0.03 pela presença da sílaba ‘ten’ nas palavras ‘extenso’ e ‘entende’.

O Poema 8 (Tabela 13) foi gerado a partir de sentenças completas de *Serafim Ponte Grande*. A estrutura foi de 2 quartetos com rima mista, com escolhas aleatórias de sílaba final, produzindo um padrão AABA ABBA. Para filtragem de sentenças candidatas, não foi estabelecida qualquer restrição métrica, gerando um poema heterométrico: 3 dodecassílabos, 1 decassílabo, 2 eneassílabos e 2 heptassílabos. Já que a

obra possui o menor número total de sentenças, todas as métricas poderiam ser consideradas. Encontravam-se, disponíveis, uma quantidade razoável de sentenças candidatas, com 27 sentenças completas terminadas em ‘te’ e 28 em ‘a’. A pontuação final foi 2.19 pontos, uma pontuação razoável principalmente porque se trata de um poema heterométrico. A acentuação grave de todas as últimas palavras da primeira estrofe, e todas as esdrúxulas na segunda, resultou numa pontuação máxima de acentuação. Isto evidencia que este é um critério de fácil pontuação, mesmo em um poema com várias métricas. Por outro lado, a pontuação pelo esquema rítmico, intermediária, foi de valor 0.61, já que a escolha pelo padrão heterométrico obteve sentenças de comprimentos diversos diminuindo as possibilidades de coincidência de posições entre as tônicas. Houve apenas uma coincidência de sílaba idêntica na mesma posição tônica, ‘men’ (‘inclemente’ e ‘comovidamente’), com pontuação limitada a 0.03. Já na avaliação fonética, foram encontrados letras iguais para além das sílabas finais em pares de versos, gerando uma pontuação 0.55, com quatro letras finais idênticas (‘ente’) entre ‘inclemente’ e ‘frente’ e entre ‘frente’ e ‘comovidamente’, além das duas ‘ia’ entre ‘platéia’ e ‘obrigatória’ e três letras ‘ite’ entre ‘leite’ e ‘noite’.

Os resultados apresentados exemplificam como poemas gerados a partir da prosa literária recebem avaliações finais variadas. O tamanho da obra fonte, com maior número de sentenças candidatas, pode favorecer a otimização

A noite desce qual um pássaro inclemente
O meu relógio anda sempre para a frente
Mostrou-nos uma carta e uma fotografia
A nova se espalha comovidamente

Amélia - Minha ama de leite
Rosáceas sobre aspargos da platéia
Vacina Obrigatória
Lalá passou mal a noite

escansão	métrica	tônicas
A/ n#oi/te/ d#es/ce/ qu#al/ um/ p#á/ssa/ro in/cle/m#en/te	Dodecassílabo	2 4 6 8 12
O/ m#eu/ re/l#ó/gi/o #an/da/ s#em/pre/ para a/ fr#en/te	Hendecassílabo	2 4 6 8 11
Mos/tr#ou/nos/ #u/ma/ c#ar/ta e #u/ma/ fo/to/gra/f#i/a	Dodecassílabo	2 4 6 7 12
A/ n#o/va/ se/ es/p#a/lha/ co/mo/vi/da/m#en/te	Dodecassílabo	2 6 12
A/m#é/li/a -/ M#i/nha/ #a/ma/ de/ l#e/i/te	Decassílabo	2 5 7 10
Ro/s#á/ce/as/ s#o/bre as/p#ar/gos/ da/ pla/t#é/i/a	Hendecassílabo	2 5 7 11
Va/c#i/na/ O/bri/ga/t#ó/ri/a	Heptassílabo	2 7
La/l#á/ pa/ss#ou/ m#al/ a/ n#o/i/te	Heptassílabo	2 4 5 7

Tabela 13: Poema 8 gerado pelo PROPOE a partir de *Grande Sertão: Veredas*

rítmica, embora não em todos casos e nem em todos critérios individuais considerados na avaliação. Os poemas exemplificados seguiram diferentes parâmetros de filtragem e de montagem, com quantidades maiores ou menores de estrofes. Mesmo com estas variações, baseadas na existência de uma quantidade suficiente de sentenças candidatas, as pontuações obtidas na avaliação rítmica exibiram mudanças pequenas de valores. Por outro lado, quando há uma quantidade limitada de sentenças, as pontuações tendem a exibir mudanças maiores de valores.

5. Considerações Finais

PROPOE é o primeiro sistema computacional a gerar poemas a partir de sentenças versificadas extraídas da prosa literária em língua portuguesa. A montagem dos poemas baseia-se na otimização combinatória de versos similares, por meio de um algoritmo guloso, que realiza a seleção de cada verso de acordo com critérios rítmicos e fonológicos. Tal estratégia permite reduzir o custo computacional, através da comparação de sentenças candidatas, gerando poemas com vários tipos de rima. Os poemas, de versos isométricos ou heterométricos, são avaliados pelos parâmetros que os caracterizam como poemas versificados. Os resultados mostram grande variação morfológica na construção dos poemas, provenientes de uma mesma obra, processada pelo MIVES. É importante enfatizar que a associação sequencial das sentenças (versos) não sofre

qualquer restrição semântica; isso é, a sucessão dos versos não se submete a critérios semânticos de organização. Entretanto, a geração de poemas, a partir da mesma obra em prosa, fornece um contexto e uma certa coesão temática.

Outro aspecto importante é que, diferente do que observamos em muitos sistemas, as sentenças são integralmente preservadas e tratadas como unidades candidatas. As sentenças são mantidas, em suas formas originais, preservando os paralelismos estruturais, especialmente fonológicos e rítmicos, encontrados na obra em prosa de que foram extraídos. Ainda sobre a organização, são desconsideradas quaisquer descontinuidades sintáticas entre os versos (enjambment). Tanto os versos resultantes, quanto as estrofes, exibem uma estrutura sem ordenamento sintático. Os poemas gerados são submetidos a avaliação do próprio sistema, que baseia-se (i) na coincidência entre letras das rimas externas, (ii) sílabas tônicas e seus posicionamentos, e (iii) esquema rítmico de pares de versos. Em muitos casos os poemas recebem pontuação nula em certos critérios, resultado, por exemplo, da ausência de coincidência entre sílabas tônicas dos esquemas rítmicos, pontuando em outros. Embora, como já afirmamos, a avaliação desconsidere aspectos semânticos, concentrando-se em propriedades estruturais (especialmente padrões métricos e rítmico-fonológicos), encontramos alguma coesão temática nas construções, resultado, como afirmamos acima, de extrações da mesma obra de origem, em todos os casos.

Novos desenvolvimentos estão planejados para o PROPOE, principalmente ampliação e flexibilização de critérios para montagem dos poemas. Ao pontuar sentenças candidatas, além do esquema rítmico, consideramos a inclusão de critérios para possíveis rimas toantes e consoantes, rima interna, assim como similaridade semântica, em diversas escalas, parágrafos, sentenças ou palavras. Baseados em múltiplos critérios, tais padrões podem ser ponderados de formas diversas pelo usuário, por meio de pesos numéricos, permitindo obter diferentes resultados na otimização da geração dos poemas.

Também planejamos ampliar ou diversificar as fontes de sentenças, com possibilidade de compor um corpora de sentenças versificadas oriundas de mais de uma obra, por exemplo, do mesmo autor, do mesmo movimento literário, e mesmo de diferentes períodos e movimentos, produzindo maiores variações e novas possibilidades para geração de poemas.

Referências


- Aamodt, Agnar & Enric Plaza. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications* 7(1). 39–59. [doi 10.3233/AIC-1994-7104](https://doi.org/10.3233/AIC-1994-7104).
- Agirrezabal, Manex, Bertol Arrieta, Aitzol Astigarraga & Mans Hulden. 2013. POS-tag based poetry generation with wordnet. Em *14th European Workshop on Natural Language Generation*, 162–166.
- Ali, Manuel Said. 1999. *Versificação Portuguesa*. EDUSP.
- Almeida, Guilherme de. 1946. A poesia d’Os Sertões. *Diário de São Paulo* 18.
- Andrade, Mário de. 1988. *Macunaíma: o herói sem nenhum caráter*. Oficinas Gráficas de Eugenio Cupolo.
- Andrade, Oswald de. 1933. *Serafim ponte grande*. Ariel, Editora Ltda.
- Antonio, Jorge Luiz. 2009. Procedimentos poéticos com o(s) computador(es). *SIBILA: Revista de poesia e crítica literária* 21.
- Balestrini, N. 1962. Tape Mark I. Em *Almanacco Letterario Bompiani*, V. Bompiani e C.
- Bena, Brendan & Jugal Kalita. 2019. Introducing aspects of creativity in automatic poetry generation. Em *16th International Conference on Natural Language Processing*, 26–35.
- Boden, Margaret A. 2009. Computer models of creativity. *AI Magazine* 30(3). 23–23. [doi 10.1609/aimag.v30i3.2254](https://doi.org/10.1609/aimag.v30i3.2254).
- Brailsford, Sally C, Chris N Potts & Barbara M Smith. 1999. Constraint satisfaction problems: Algorithms and applications. *European Journal of Operational Research* 119(3). 557–581. [doi 10.1016/S0377-2217\(98\)00364-6](https://doi.org/10.1016/S0377-2217(98)00364-6).
- de Campos, Augusto. 1996. Transertões. *Folha de São Paulo* 3 de novembro de 1996.
- Carvalho, Cid Ivan da Costa, Davis Macedo Vasconcelos & Leonel Figueiredo de Alencar Araripe. 2012. Superando o estado da arte na etiquetagem morfossintática por meio de regras de pós-etiquetagem. Em *X Encontro de Linguística de Corpus: Aspectos metodológicos dos estudos de corpora*, 122–134.
- Carvalho, Ricardo, Angelo Loula & João Queiroz. 2020. Identificação computacional de estruturas métricas de versificação na prosa de euclides da cunha / computational identification of versification metric structures in euclides da cunha’s prose. *Revista de Estudos da Linguagem* 28(1). 41–68.
- Castro, Leandro N. 2006. *Fundamentals of natural computing: basic concepts, algorithms, and applications*. Chapman and Hall/CRC.
- Colton, Simon & Geraint A. Wiggins. 2012. Computational creativity: The final frontier? Em *20th European Conference on Artificial Intelligence*, 21–26.
- Cormen, Thomas H, Charles E Leiserson, Ronald L Rivest & Clifford Stein. 2009. *Introduction to algorithms*. MIT press.
- Cunha, Euclides da. 1902. *Os Sertões*. Fundação Biblioteca Nacional.
- D’Ambrosio, Matteo. 2018. The early computer poetry and concrete poetry. *MATLIT: Materialidades da Literatura* 6. 51–72.
- Das, Amitava & Björn Gambäck. 2014. Poetic machine: Computational creativity for automatic poetry generation in bengali. Em *5th International Conference on Computational Creativity (ICCC)*, 230–238.
- Durkin, John. 1994. *Expert Systems: Design and development*. Prentice Hall PTR.
- Eiben, Agoston E. & James E. Smith. 2003. *Introduction to evolutionary computing*. Springer. [doi 10.1007/978-3-662-05094-1](https://doi.org/10.1007/978-3-662-05094-1).
- Gatt, A. & E. Krahmer. 2018. Survey of the state of the art in natural language generation: Core

- tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61. 65–170.
- Gervás, Pablo. 2000. WASP: Evaluation of different strategies for the automatic generation of spanish verse. Em *AISB-00 Symposium on Creative & Cultural Aspects of AI*, 93–100.
- Gervás, Pablo. 2001. An expert system for the composition of formal spanish poetry. Em *Applications and Innovations in Intelligent Systems VIII*, 19–32. Springer.
- Gervás, Pablo. 2016. Constrained creation of poetic forms during theme-driven exploration of a domain defined by an n-gram model. *Connection Science* 28(2). 111–130. doi 10.1080/09540091.2015.1130024.
- Goldstein, Norma Seltzer. 2008. *Versos, sons, ritmos*. Editora Ática.
- Gonçalo Oliveira, Hugo. 2012. PoeTryMe: a versatile platform for poetry generation. *Computational Creativity, Concept Invention, and General Intelligence* 1. 21.
- Gonçalo Oliveira, Hugo. 2017a. O poeta artificial 2.0: Increasing meaningfulness in a poetry generation twitter bot. Em *Workshop on Computational Creativity in Natural Language Generation (CC-NLG)*, 11–20.
- Gonçalo Oliveira, Hugo. 2017b. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. Em *10th International Conference on Natural Language Generation*, 11–20. doi 10.18653/v1/W17-3502.
- Gonçalo Oliveira, Hugo, F. Amilcar Cardoso & Francisco C. Pereira. 2007. Tra-la-Lyrics: An approach to generate text based on rhythm. Em *4th International Joint Workshop on Computational Creativity*, s/pp.
- Hämäläinen, Mika, Khalid Alnajjar & Thierry Poibeau. 2022. Modern french poetry generation with roberta and gpt-2. Em *International Conference on Computational Creativity (ICCC)*, s/pp.
- Jakobson, Roman & Krystyna Pomorska. 1985. *Diálogos*. Editora Cultrix.
- Kay, Martin. 1996. Chart generation. Em *34th Annual Meeting on Association for Computational Linguistics*, 200–204. doi 10.3115/981863.981890.
- Lutz, Theo. 1959. Stochastische texte. *augenblick* 4(1). 3–9.
- Manurung, Hisar. 2004. *An evolutionary algorithm approach to poetry generation*: University of Edinburgh. College of Science and Engineering. Tese de Doutoramento.
- Manurung, Hisar, Graeme Ritchie & Henry Thompson. 2000. Towards a computational model of poetry generation. Relatório técnico. The University of Edinburgh.
- Netzer, Yael, David Gabay, Yoav Goldberg & Michael Elhadad. 2009. Gaiku: Generating Haiku with word associations norms. Em *Workshop on Computational Approaches to Linguistic Creativity*, 32–39.
- Otter, Daniel W., Julian R Medina & Jugal K. Kalita. 2020. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems* 32(2). 604–624. doi 10.1109/TNNLS.2020.2979670.
- Perloff, Marjorie. 2013. *O gênio não original: poesia por outros meios no novo século*. UFMG.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei & Ilya Sutskever. 2019. Language models are unsupervised multitask learners. openai blog. 2019; 1 (8). https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
- Rahman, Fahrurrozi & Ruli Manurung. 2011. Multiobjective optimization for meaningful metrical poetry. Em *International Conference on Computational Creativity (ICCC)*, 4–9.
- Rashel, Fam & Ruli Manurung. 2014. Pemuisi: a constraint satisfaction-based generator of topical indonesian poetry. Em *International Conference on Computational Creativity (ICCC)*, 82–90.
- Rosa, João Guimarães. 1956. *Grande sertão: veredas*. Livraria José Olympio Editora.
- Ryan, Marie-Laure, Lori Emerson & Benjamin J Robertson. 2014. *The Johns Hopkins guide to digital media*. JHU Press.
- Silva, Pedro Paulo da (ed.). 2014. *Teoria da Literatura I*. Editora Pearson.
- Sousa, Tatiane. 2007. *Haicais de Bashô: O oriente traduzido no ocidente*: Universidade Estadual do Ceará, Fortaleza. Tese de Mestrado.
- Souza, Erthos Albino de. 1991. Le Tombeau de Mallarmé. Em *Mallarmé*, Perspectiva.
- Spina, Segismundo. 2003. *Manual de versificação românica medieval*. Ateliê Editorial.

Tobing, Bertly Chrismartin Lumban & Ruli Manurung. 2015. A chart generation system for topical metrical poetry. Em *International Conference on Computational Creativity (ICCC)*, 308–314.

Wong, Martin Tsan, Andy Hon, Wai Chun & Tat Chee. 2008. Automatic haiku generation using VSM. Em *7th WSEAS International Conference on Applied Computer and Applied Computational Science*, 318–323.

Zugarini, Andrea, Stefano Melacci & Marco Maggini. 2019. Neural poetry: Learning to generate poems using syllables. Em *International Conference on Artificial Neural Networks*, 313–325.

 [10.1007/978-3-030-30490-4_26](https://doi.org/10.1007/978-3-030-30490-4_26).