



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

Ricardo Carvalho

Universidade Estadual de Feira de Santana (UEFS), Feira de Santana, Bahia / Brasil
ricardo.sys@gmail.com

<https://orcid.org/0000-0001-7785-562X>

Angelo Loula

Universidade Estadual de Feira de Santana (UEFS), Feira de Santana, Bahia / Brasil
angelocl@uefs.br

<https://orcid.org/0000-0001-7802-1731>

João Queiroz

Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Minas Gerais / Brasil
queirozj@gmail.com, queirozj@pq.cnpq.br

<https://orcid.org/0000-0001-6978-4446>

Resumo: Estruturas de versificação em prosa de língua portuguesa constituem um fenômeno ainda inexplorado por teóricos e historiadores da literatura, e a mineração automática de tais estruturas é inédita em Linguística Computacional. O sistema MIVES (*Mining Verse Structure*) foi desenvolvido para escansão computacional de estruturas métricas de versificação em prosa de língua portuguesa. Ele é capaz de identificar, classificar e comparar, frequência, densidade e dispersão de estruturas heterométricas de versificação, distribuídas em diversas escalas de observação -- de uma obra ou autor, até períodos e movimentos literários. Apresentamos o sistema, e fazemos uma validação preliminar dele em três obras de Euclides da Cunha (*Os Sertões*, *À Margem da História*, *Contrastes e Confrontos*). Elas foram selecionadas porque constituem o corpus principal de um dos mais importantes prosadores de língua portuguesa e porque *Os Sertões* foi objeto do que Augusto de Campos chamou de “leitura verso-espectral”, uma operação capaz de revelar “mais de 500 decassílabos na obra”. MIVES identificou estruturas métricas em 48,18% das sentenças de *Os Sertões*, 48,96% de *À Margem da História* e

37,48% de *Contrastes e Confrontos*, uma taxa surpreendente, quando comparada aos resultados exibidos por Augusto de Campos. Pode-se afirmar que MIVES inaugura novos padrões de observação e análise de estruturas ainda não investigadas na prosa literária de língua portuguesa.

Palavras-Chave: estruturas de versificação; prosa; mineração de versos; Euclides da Cunha.

Abstract: Versification structures written in Portuguese language prose are still an unexplored phenomenon to literary theorists and historians, and the automatic mining of such structures is still unseen in Computational Linguistics. The MIVES (Mining Verse Structure) system was developed for computational scansion of metric versification structures in Portuguese language prose. It is able to identify, classify and compare, frequency, density and dispersion of heterometric structures of versification, distributed at different scales of observation, from a work or author, to historical periods and movements. We present the system, and a preliminary validation analysing three works by Euclides da Cunha (*Os Sertões*, *À Margem da História*, *Contrastes e Confrontos*). They were selected because they constitute the main corpus of one of the most important Portuguese-language writers, and because *Os Sertões* was the object of what Augusto de Campos (2010, p. 14) called “verse-spectral reading”, an operation able of revealing “more than 500 decasyllables in the book, among sapphic and heroic verses, and more than two hundred dodecasyllables. MIVES identified metric structures in 48.18% of the sentences of *Os Sertões*, 48.96% of *À Margem da História* and 37.48% of *Contrastes e Confrontos*, an unexpected rate when compared to the partial results of Augusto de Campos. It can be said that MIVES inaugurates new patterns of observation and analysis of structures not yet investigated in Portuguese language literary prose.

Keywords: structures of versification; verse mining; prose; Euclides da Cunha.

Recebido em 28 de fevereiro de 2019

Aceito em 05 de julho de 2019

1. Introdução

MIVES (*Mining Verse Structure*)¹ foi desenvolvido para escansão computacional de estruturas métricas de versificação em prosa de

¹ O sistema está disponível em <http://sites.ecomp.uefs.br/lasic/projetos/mives>. Ele resulta da colaboração entre o Laboratório de Sistemas Inteligentes e Cognitivos (LASIC-UEFS), coordenado por Angelo Loula, e o Iconicity Research Group (IRG, UFJF), coordenado por João Queiroz. Este sistema está relacionado à dissertação de mestrado de Carvalho (2017).

língua portuguesa. Trata-se de um fenômeno inexplorado por teóricos e historiadores da literatura, e inédito em Linguística Computacional. Uma validação preliminar do sistema inclui a análise de três obras de Euclides da Cunha – *Os Sertões*, *A Margem da História* e *Contrastes e Confrontos*. Elas foram selecionadas porque constituem o *corpus* principal de um dos mais importantes prosadores de língua portuguesa, e porque os *Sertões* foram objeto do que Augusto de Campos chamou de “leitura verso-espectral” (CAMPOS, 2010, p. 14), uma “operação manual”² de mineração de estruturas de versificação “mal escondidas” sob a prosa de Euclides, capaz de revelar “mais de 500 decassílabos no livro, entre sáficos e heróicos, e mais de duas centenas de dodecassílabos” (CAMPOS, 2010, p. 14).

É bem conhecida a tese de que a “estrutura da poesia é um paralelismo contínuo” (JAKOBSON, 1988, p. 102), em vários níveis de descrição, linguísticos e paralinguísticos: paralelismos gramaticais, sintáticos, fonológicos, rítmicos, e até tipográficos. Este fenômeno não tem correspondência na prosa. Mas, como afirma Jakobson (1988, p. 106), a prosa literária “ocupa um lugar intermediário entre a poesia enquanto tal e a língua de comunicação comum, prática, não se devendo esquecer que é incomparavelmente mais difícil analisar um fenômeno intermediário, de transição, do que estudar fenômenos extremos”. Neste domínio, diversas formas de paralelismo se distribuem em muitos níveis de descrição, permitindo inclusive a identificação de estruturas métricas de versificação sob a superfície aparentemente mais uniforme da prosa, um fenômeno surpreendente e ainda inexplorado.

O *verso* é usualmente definido como a linha do poema. Ele pode ter diversas medidas ou extensões, e pode se organizar de acordo com diversos padrões. O metro é a medida do verso (SPINA, 2003, p. 29). Em português, o sistema de metrificação é silábico-acental -- conta-se o número de sílabas de cada verso e verifica-se a alternância entre sílabas fortes e fracas. A alternância regular cria certos padrões que, combinados às repetições posicionais das sílabas, cria segmentos internos, estabelecendo as regras de versificação ou metrificação (ALI, 2006). O processo de identificação e classificação dos padrões historicamente normatizados

² Trata-se, evidentemente, de uma “simplificação”, para efeito classificatório, e em oposição à operação computacional, chamar esta operação (escansão de versos) de “manual”. Ela resulta de uma complexa interação entre “percepção categorial” (percepção de classes e categorias) e diversas formas de inferência analógica e indutiva.

pela tradição literária é chamado de *escansão*. Não decorre deste processo, que atua predominantemente em poemas de versos metrificados, um resultado inequívoco, de classificação dos padrões, livre de contexto. Um verso pode, porque possui certos encontros vocálicos, ser escandido como um decassílabo (dez sílabas poéticas) ou um eneassílabo (nove sílabas poéticas). A decisão sobre como deve comportar-se depende, usualmente, do contexto, dos versos antecedentes, e da tradição histórica.

A prosa não constitui objeto da *escansão*, em qualquer de suas modalidades, inclusive literária. Entretanto, alguns estudiosos se detiveram na identificação de estruturas métricas de versificação na prosa. Em língua portuguesa, são pioneiros os trabalhos de Guilherme de Almeida (1946) e Augusto de Campos (2010). Campos, em *A Poética de Os Sertões*, publicado em 1997 (reeditado em 2010), retoma e amplia o “projeto de prospecção” de Almeida, publicado originalmente em 1946, e revela estruturas deca e dodecassilábicas “mal escondidas” sob a prosa de Euclides da Cunha. Para Campos (2010, p. 297-298), “[Euclides da Cunha], a propósito, escrevia poesia e conhecia bem a métrica, embora nada tivesse escrito de relevante como poeta. Ao constatar esses padrões rítmicos definidos na sua obra em prosa, achei que seria interessante e útil anotá-los e acentuá-los.” Se tais estruturas resultam de uma experimentação deliberada, se são “fruto de pura intuição ou de consciente artesanaria” (CAMPOS, 2010, p. 27), isso pouco importa no domínio de descoberta empírica do fenômeno. O fato mais relevante é que há, na prosa euclidiana, estruturas heterométricas de versificação, de padrões rítmicos variados. Pode-se perguntar: este fenômeno pode ser generalizado para toda obra de Euclides? Sua frequência (e distribuição) é regular ao longo de suas obras? Há estruturas métricas de versificação na prosa literária de outros autores, de outros períodos?

Diversos sistemas desenvolvidos recentemente em Linguística Computacional concentram-se na automatização da *escansão* de poemas em versos metrificados (MITTMANN, WANGENHEIM, SANTOS, 2018; PLAMONDON, 2006; ARAÚJO; MAMEDE, 2002; GERVÁS, 2000). Mas Oostendorp (2014), por exemplo, desenvolveu um sistema que minera padrões de alternância de tônicas na prosa. Nosso propósito assemelha-se ao de Oostendorp, embora seja mais geral e ambicioso. MIVES está relacionado à concepção de um modelo computacional sincrônico-diacrônico de análise estatística de distribuição de padrões heterométricos de versificação na prosa, relação com o período literário correspondente, com as obras de um autor, de diversos autores, e de diversos períodos literários. O que vamos apresentar aqui é apenas a

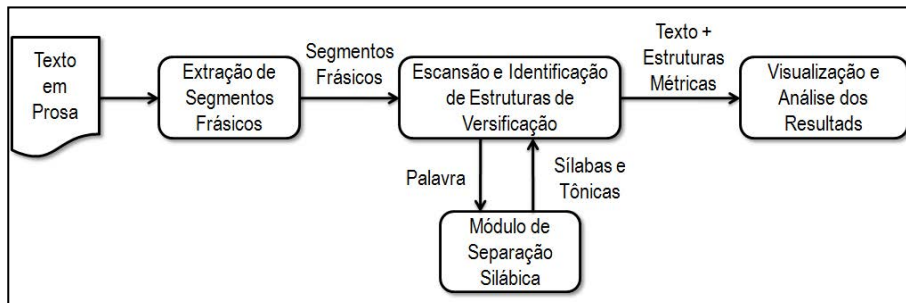
primeira etapa deste projeto. Nosso propósito é analisar a frequência e a densidade de estruturas métricas em *Os Sertões*, de Euclides da Cunha, e comparar os resultados com análises de *A Margem da História e Contrastes e Confrontos*, do mesmo autor.

2. Identificação automática de estruturas métricas de versificação

MIVES (*Mining Verse Structure*) foi desenvolvido, como afirmamos, para identificar e classificar, na prosa literária, sentenças ou trechos de sentenças que possuem estruturas métricas de versificação. O sistema extrai as sentenças do texto, identifica e classifica estruturas métricas procuradas pelo usuário, e fornece uma visualização dos resultados obtidos. O maior desafio encontra-se no processo de identificação de estruturas métricas, uma vez que não decorre da escansão um resultado único, inequívoco, independente de contexto. Como trata-se de prosa, não há uma clara demarcação do início e do fim das estruturas, como encontramos mais facilmente nos versos de um poema metrificado. Na prosa, a estrutura métrica pode ser formada por uma sentença completa, ou por um trecho de sentença.

O processamento (Figura 1) tem início com a extração de sentenças de um arquivo de texto. Cada sentença, em seguida, é segmentada em palavras para separação silábica e identificação de sílabas tônicas. Depois, ela é submetida à escansão, que considera variações normativamente aceitas de separação silábica. Determina-se, então, se a sentença, ou um trecho dela, possui um padrão métrico, e são indicadas possíveis alternativas. O texto, cujas sentenças métricas e variações foram identificadas, é enviado para uma interface de visualização, navegação e análise dos resultados.

FIGURA 1 – Etapas de processamento do MIVES



Extração das Sentenças

A partir da cópia digital de um texto em prosa, é feita uma segmentação em sentenças – um segmento frásico na forma de uma sequência de palavras que termina em um marcador. Chamamos este marcador de “ponto final da sentença”. Ele pode ser o ponto de um segmento, ou um ponto final, funcionando como um delimitador. O principal desafio desta etapa é identificar pontos que não são delimitadores, como aqueles utilizados nos pronomes de tratamento. Além disso, é necessário considerar casos como títulos e subtítulos, que geralmente não terminam com um delimitador e devem ser tratados como sentenças. As fases posteriores de processamento têm foco nas sentenças, na busca de sentenças completas, ou trechos de sentenças.

Separação silábica e marcação de sílabas tônicas

Embora os protocolos de separação de sílabas poéticas e gramaticais não produzam, necessariamente, resultados coincidentes, a separação gramatical de sílabas é um passo inicial para realização da separação poética. Nesta etapa, uma sentença como “Hipóteses sobre a sua gênese.” pode ser escandida como “Hi/p#ó/te/ses/ s#o/bre a #u/a g#ê/ne/se”, onde / indica um separador de sílaba e # um marcador de sílaba tônica. A separação silábica realizada pelo sistema, e a marcação das tônicas, utilizam um algoritmo desenvolvido por Neto, Rocha e Souza (2015), associado a um dicionário interno do MIVES, para obter melhor desempenho e para tratar exceções e falhas do sistema de Neto, Rocha e Souza (2015). O algoritmo de Neto, Rocha e Souza (2015) baseia-se em 20 regras fonológicas propostas inicialmente por Silva, Braga e Resende Jr (2008). A ideia aqui é que todas as sílabas possuem uma vogal como núcleo, que pode ser cercada por consoantes ou por outras vogais (semi-vogais ou glide). Segundo Neto, Rocha e Souza (2015), o algoritmo foi testado em um extrato de 10.000 palavras escolhidas aleatoriamente na base de dados do CETENFolha, obtendo uma taxa de acerto de 99.14%. Como afirmamos, além deste algoritmo, o MIVES usa um dicionário com palavras associadas a uma versão com separação silábica gramatical.

O processamento de separação silábica é lento. Assim, optamos por armazenar palavras separadas silabicamente em uma estrutura de dados de dicionário baseado em mapa,³ em que as chaves são organizadas

³ Um mapa é uma estrutura de dados que mantém uma coleção de chaves associadas a um valor, em pares (chave, valor). Nesta estrutura, uma chave ocorre uma única vez.

por tabela *hash*,⁴ e são armazenadas em disco para persistência.⁵ Para obter a separação silábica, e marcação de tônicas, realiza-se uma busca no dicionário. Caso a palavra não seja encontrada, é executado o algoritmo de Neto, Rocha e Souza (2015) e o resultado é armazenado para consultas. Além de otimizar o tempo de processamento, o dicionário é utilizado para armazenar palavras que não são processadas adequadamente pelo algoritmo de Neto, Rocha e Souza (2015). Por exemplo, “euforia” deveria ser separado como “eu/fo/ri/a” e não como “eu/fo/ria”, resultado obtido pelo algoritmo. Diversas exceções foram incluídas no dicionário do MIVES e novas palavras podem ser incluídas pelo usuário.

Escansão e identificação das estruturas métricas de versificação

Mas a escansão não termina na fase de separação silábica das palavras, consideradas isoladamente, e marcação das tônicas. Fenômenos vocálicos intervocabulares (entre as palavras), ou intravocabulares (no interior da palavra), são considerados, determinando diversas variações aceitas. Em certos casos, os resultados de escanções podem coincidir com resultados da separação silábica gramatical, mas não se trata de uma norma. Esta etapa da escansão constitui o núcleo do MIVES. São considerados fenômenos intravocabulares (crase, sinérese, diérese) e intervocabulares (crase, sinalefa, elisão). Eles cobrem um grande espectro de ocorrências. Outros fenômenos podem afetar a escansão, como aférese, apócope ou síncope, mas eles são raros e de difícil especificação.

A crase intravocabular é caracterizada pela união, em uma sílaba, de vogais idênticas tratadas originalmente como hiato; por exemplo, “saara” (“sa/a/ra”) transforma-se em “saa/ra”. A sinérese é caracterizada pela transformação de hiatos em ditongos; por exemplo, “magoado” (“ma/go/a/do”) torna-se “ma/goa/do”. Menos frequente, a diérese é caracterizada pela transformação de um ditongo em um hiato; por exemplo, “saudade” (“sau/da/de”) torna-se “sa/u/da/de”. A crase intervocabular é a fusão de duas vogais idênticas, ao final de uma palavra e início da palavra subsequente; por exemplo, no verso “É/ a es/car/pa a/brup/ta e/ vi/va/ dos/ pla/nal/tos.”,

⁴ Também conhecida como “tabela de espalhamento”. É uma estrutura de dados que utiliza chaves associadas a valores e permite consulta rápida de um valor a partir de uma chave.

⁵ Relaciona-se ao armazenamento de dados de forma não-volátil, permitindo que, em qualquer momento, uma informação seja recuperada a partir de um dispositivo de processamento de dados autorizados.

com crase em “escarpa abrupta”. A sinalefa é caracterizada pela passagem da primeira vogal para semivogal, formando um ditongo com a vogal da palavra subsequente; por exemplo, no verso “A/cu/do eu/ lo/go, e, en/ quan/to o/ re/mo a/per/to.”, o fenômeno aparece entre a última sílaba de “Acudo” e “eu”. A elisão é caracterizada pela supressão da primeira vogal em detrimento da seguinte; por exemplo, entre a última sílaba de “como” e “uma” no verso “co/mo u/ma/ nó/doa am/plís/si/ma/ de/ san/gue.”.

O processo de escansão não produz resultados inequívocos. Os fenômenos descritos (intra e intervocabulares) podem ser considerados, ou não. A aplicação das regras é decidida pela escolha de pares de vogais a serem considerados para cada fenômeno vocálico. A busca por estruturas respeita as delimitações do usuário do sistema, que pode indicar, para cada encontro vocálico (par de sílabas), quais regras de fusão ou separação devem ser aplicadas. Além disso, o usuário decide os limites dos metros investigados (contagem total de sílabas), com valores mínimos e máximos. Durante a escansão, as escolhas das regras vocálicas e os limites de metrificação indicados pelo usuário determinam as variações a serem consideradas pelo sistema.

Neste módulo, o algoritmo de escansão realiza uma análise para identificar possíveis ocorrências de crase, sinérese e diérese. Ao encontrar vogais átonas, iguais e adjacentes na palavra, caracterizando uma crase, o algoritmo é capaz de transformar, conforme instruções fornecidas pelo usuário, duas sílabas em uma. Os casos de sinérese e diérese também dependem de um conjunto de parâmetros fornecidos pelo usuário. Ele informa que combinações de vogais adjacentes devem ser consideradas, e como estas combinações devem ser utilizadas quando identificadas no interior de uma palavra. Para a sinérese e diérese, o usuário deve informar pares de vogais, que serão tratados como encontros vocálicos ou ditongos e, para cada par informado, qual deve ser o comportamento do sistema durante a escansão. Nesta etapa, “É mais um inimigo a supplantar.” torna-se, por exemplo, “É/ mais/ um/ i/ni/mi/go a/ su/plan/tar.”. Como não é produzido um resultado inequívoco, o mesmo verso pode ser escandido diferentemente. Assim, na busca por estruturas entre dez e doze sílabas métricas, e aceitando variações previstas, uma sentença escandida como “Es/ta i/lu/são/ é/ em/pol/gan/te ao/ lon/ge.” (10 sílabas) torna-se “Es/ta/ i/lu/são/ é/ em/pol/gan/te ao/ lon/ge.” (11 sílabas), ou “Es/ta/ i/lu/são/ é/ em/pol/gan/te/ ao/ lon/ge.” (12 sílabas).

A busca por estruturas métricas não está restrita a sentenças completas. Trechos iniciais ou finais de sentenças podem ser avaliados, conforme decisão do usuário. Para um trecho de início, ou final, de sentença, sinais de pontuação tais como ponto, vírgula, ponto-vírgula, dois pontos, reticências e exclamação, são utilizados como delimitadores. Para a identificação de trechos, as sentenças são escandidas até que se encontre uma estrutura métrica adequada aos padrões designados pelo usuário. Só então é verificada a existência de um delimitador para validá-la como um trecho aceitável. Assim, trechos iniciais de sentenças escandidas são considerados válidos se terminados com um dos sinais de pontuação mencionados. Depois de escandidas as sentenças, ou trechos de sentenças, são identificadas aquelas que satisfazem as estruturas métricas designadas pelo usuário. As versões escandidas são associadas às sentenças originais no texto, e o texto processado é enviado para visualização e análise.

Visualização e análise de resultados

Para avaliação dos resultados, foi desenvolvida uma interface para navegação e análise das escansões, com múltiplas visões dos resultados (Figuras 2 e 3). As sentenças são visualmente destacadas no texto original. Isto permite ao usuário avaliar a sentença identificada, comparando-a com sentenças próximas. Para uma visualização somente das sentenças identificadas, é organizada uma lista na ordem de ocorrência do texto, e suas escansões.

Além da identificação das sentenças, o sistema gera gráficos que permitem visualizar a dispersão das sentenças ao longo do texto processado e a frequência absoluta dos padrões métricos classificados em ordem decrescente (do mais ao menos frequente). Pode-se também obter gráficos que representam as distâncias, ou intervalos, entre as sentenças. Para sentenças que não foram identificadas como métricas, a interface permite que o usuário verifique como foi realizada a escansão, de modo que ele pode avaliar o motivo da exclusão de um determinado segmento dos resultados. Para analisar os resultados através de outras ferramentas, pode-se exportar as sentenças encontradas, escandidas e classificadas, e a relação de todas as sentenças (métricas ou não) com a indicação de metros escandidos.

FIGURA 2 – Interface do sistema

Mining Verse Structure - MIVES V 1.0 - Beta

Analise Exportar

Entre com o texto para realizar

DECASSÍLABO

HENEDECASSÍLABO

DODECASSÍLABO

VERSOS ENCONTRADOS

ESTATÍSTICAS

TEXTO PROCESSADO

Capítulo BO sertanejo. **O sertanejo é, antes de tudo, um fofo.** Não tem o requizito exaltivo dos mestiçosneurastênicos do literal. A sua aparência, entretanto, ao primeiro lance de vista, revela o contrário. Faltalhe solística inspecível, o desempenho, a estrutura corretíssima das organizações atléticas. **É desgracioso, desimprovisado, fofo.** Hércules-Quasimodo, reflete no aspecto a fealdade tipicados fracos. O andar sem firmeza, sem aprumo, quase gigante e sinuoso, aparenta a transição demembros desarticulados. Agrava-o a postura normalmente abduzida, num manifestar de dissidênciaque lhe dá um caráter de humildade deprimente. A pé, quando parado, recostase-lhe invariavelmente no primeiro umbral ou parede que encontrar o cavalo, se suporta o animal para brincar duas palavras comum conhecido, cal logo sobre um dos estribos, descansando sobre a espinha da sela. Caminhando, mesmo a passo rápido, não traga trajetória retilínea e firme. Avança calientemente, num bambolearcaracterístico, de que parecem ser o traço geométrico os meandros das trilhas sertanejas. E se namarcha estaca pelo motivo mais vulgar, para enrolar um cigarro, bater o loqueiro, ou travar ligeiraconversa com um amigo, cal logo — cal é o termo — de rícoras, atravessando largo tempo numaposição de equilíbrio instável, em que todo o seu corpo fica suspenso pelos dedos grandes dos pés, sentado sobre os calcaneares, com uma simplicidade e um tempo ridícula e adorável.

É o homem permanentemente fofofofo. Reflete a preguiça invencível, a stonia muscular perene, em tudo: na palavra remorada, no gesticontrafeito, no andar desagradado, na cadência languida das modélias, na tendênciaconstante à imobilidade e à quietude. Entretanto, toda esta aparência de canoço fofo. Nada é mais surpreendedor do que vê-la desaparecer de improviso. Naquelaorganização combatida operam-se, em segundos, transmutações completas. Basta aparecimento de qualquer incidente exigindo-lhe o desencadear das energias adormecidas. O homem tranfigura-se. Empertiga-se, estadeando novos relevos, novas linhas na estatura no gesto; e a cabeça firma-se-lhe, alta, sobre os ombros possantes aclarada pelo olhardesacombrado e forte; e contigem-se-lhe, prestes, numa descarga nervosa instantânea, todosos efeitos do relaxamento habitual dos órgãos; e de figura vulgar do tabaréu canestronoponta, inesperadamente, o aspecto dominador de um tãã acobreado e potente, numdesdobramento surpreendente de força e agilidade extraordinária. **Este contraste impõe-se ao mais leve exame.** Revela-se a todo o momento, em todos ospormentres da vida sertaneja — caracterizado sempre pela intercadência impressionadorentre extremos impulsos e apatias longas. É impossível idear-se cavaleiro mais churco e desleigante; sem posição, pernas coladasao bojo da montaria, tronco pendido para a frente e oscilando à feição da andadura dospequenos cavalos do sertão, desferendos e maltratados, resistentes e rápidos como poucos. Nesta atitude indolente, acompanhando mousamente, a passo, pelas chapadas, o passo tardo

-65-

Página: => 65

das boiadas, o vaqueiro preguiçoso quase transforma o "campeão" que cavalga na redeemolcedora em que atravessa dois terços de existência. Mas se uma réis "alevantada" envereda, esquiva, adiante, pela castinga garranchenta, ou se uma ponta de pado, ao longe, se trasmalha, ei-lo em momentos transformado, traçandoos arcos de rosetas largas nas ilhanças da montaria e partindo como um dardo, abufandose velozmente nos dédolos inextricáveis das juremas. Vimo-lo neste steeple-chase bárbaro. Não há contê-lo, então, no ímpeto. Que se lhe antolhem quebradas,

Anterior Próximo

MIVES Voltar Avançar Finalizar

FIGURA 3 – Interface do sistema

Mining Verse Structure - MIVES V1.0 - Beta

Analisar | Exportar

TEXTO PROCESSADO

Entre com o texto para realizar

DECASSÍLABO

A demonstração é positiva.
 Balança e influencia do Tietê.
 Função histórica do rio S. Francisco
 O melo abraça os e guarda os.
 Ligeiros nos no espago e no tempo.
 Não tiveram um historiador.
 Esta tendência é expressiva.
 O sertanejo é, antes de tudo, um forte.
 O/ ser/ta/na/je/ra É, ran/tes/ de/ tra/ido, um/ fôr/te.
 4 5 8 10
 Decassílabo
 O vaqueiro do Norte É a sua antítese.
 Aprestou-se, cedo, para a luta.
 Calcula filamento o pugilato.
 Assim todo sertanejo É vaqueiro.
 Decore líes e vida variada e farta.
 Os vaqueiros são-lhes servos submissos.
 Sabelhes, quando muito, os nomes.
 Ferrado o boi, está garantido.
 É assim passam numa agitação estéril.
 Realizam-na de junho a julho.
 Vai com ele às últimas libocas.

HENEDECASSÍLABO

DODECASSÍLABO

MIVES

Capítulo 10 sertanejo. **O sertanejo é, antes de tudo, um forte.** Não tem o requizismo exustivo dos mestiços neurastênicos do litoral. A sua aparência, entretanto, ao primeiro lance de vista, revela o contrário. Falta-lhe a plástica impecável, o **forte**. Hércules-Quase-gigante e sinovial, manifesta de digno e invariavelmente apoiado com um conhecido, calçado rápido, não traça trajetória retílinea e firme. Avança celeremente, num bambaleio característico, de que parecem ser o traço geométrico os meandros das trilhas sertanejas. É se samarita estaca pelo motivo mais vulgar, para enrolar um cigarrto, bater o loqueiro, ou travar ligeira conversa com um amigo, cai logo — cai é o termo — de cócoras, abrevessando largo tempo numa posição de equilíbrio instável, em que todo o seu corpo fica suspenso pelos dedos grandes dos pés, sentado sobre os calcanhares, com uma simplicidade a um tempo ridícula e adorável.

O homem permanentemente fatigado. Reflete a preguiça invenível, a atonia muscular pesene, em tudo: na palavra remorada, no gesto contrafeito, no andar desaprumado, na cadência langurosa das modinhas, na tendência constante à imobilidade e à quietude. Entretanto, toda esta aparência de cansaço fude. Nada é mais surpreendente do que vê-lo desaparecer de improviso. Naquele organização combatida opem-se, em segundos, transmutações completas. Basta o aparecimento de qualquer incidente exigindo-lhe o desencadear das energias adormecidas. O homem transfigura-se. Empertiga-se, estadeando novos relevos, novas linhas na estature e no gesto; e a cabeça firma-se lhe, alta, sobre os ombros possantes acilarado pelo olhar desacombrado e forte; e corrigem-se lhe, prestes, numa descarga nervosa instantânea, todos os efeitos do relaxamento habitual dos órgãos; e da figura vulgar do tabaréu cañestronoponta, inesperadamente, o aspecto dominador de um **tsã** acobreado e potente, num desdobramento surpreendente de força e agilidade extraordinária. **Este contraste impressiona de mais forte maneira.** Revolve-se a todo o momento, em todos os pormenores da vida sertaneja — caracterizado sempre pela intercação das impressões entre extremos impulsos e apatias longas. É impossível deixar-se cavaleiro mais churo e desalegant; sem posição, pernas coladas ao bojo de montaria, tronco pendido para a frente e oscilando à feição de andadura dos pequenos cavalos do sertão, desfermados e maltratados, resistentes e rápidos como poucos. Nesta atitude indolente, acompanhando morosamente, a peso, pelas chapadas, o passo tardo

—66—

Página: >> 88

das boladas, o vaqueiro preguiçoso quase transforma o "campello" que cavalga na redemolecedora em que atravessa dois terços da existência. Mas se uma réis "alevantada" envereda, espalha, adiante, pela caatinga garranchenta, ou se uma ponta de gado, ao longe, se trasmalha, e-lo em momentos transformado, trançando aciatos de rosetas largas nas ilhangas da montaria e partindo como um dardo, abufandoo velozmente no dédelos inextricáveis das juremas. Vimo-lo neste steeple-chase bárbaro. Não há comê-lo, então, no impeto. Que se lhe antolhem quebradas,

Anterior Próximo

VERSOS ENCONTRADOS

Bar chart showing the distribution of verses found:

- Decassílabo (Green bar)
- Dodecassílabo (Yellow bar)
- Dodecassílabo (Red bar)

ESTATÍSTICAS

Voltar Avançar Finalizar

3. Resultados

Exibimos os resultados obtidos em três obras de Euclides da Cunha – *Os Sertões*, *A Margem da História* e *Contrastes e Confrontos*. Apresentamos dois protocolos de validação: uma escansão de *Os Lusíadas*, e uma comparação com os resultados obtidos por Augusto de Campos em *Os Sertões*. Para processar estas obras, MIVES foi configurado para buscar somente estruturas métricas de 10 a 12 sílabas. Quanto à parametrização dos fenômenos vocálicos, foram consideradas todas as fusões e separações vocálicas para todas combinações de vogais.

Validação da escansão

Para validar o processo de escansão do MIVES, porque não há uma fonte de comparação equivalente (versos na prosa literária) além daquela fornecida por Augusto de Campos (2010), utilizamos *Os Lusíadas*, composta de 8.816 versos decassilábicos. Como ela está estruturada em estrofes de versos, foi necessário realizar uma adequação na estrutura do texto para o sistema reconhecer cada linha do poema como um segmento frásico. Esta operação foi baseada na substituição de todos os sinais de pontuação, no final de cada verso, por um ponto de final de sentença, também inserido nos versos que não possuíam sinais de pontuação. Para validação, foram procurados versos decassilábicos. MIVES escandiu 8.256 versos, com uma taxa de acerto de cerca de 94%. Os 6% não escandidos resultam de aspectos relacionados a eclipse, erros do separador silábico gramatical, estrangeirismos, e uso de diacríticos na obra de Camões, pouco usuais hoje.

Validação da mineração em prosa

Campos (2010, p. 14) afirma ter encontrado mais de quinhentos decassílabos e mais de duas centenas de dodecassílabos em *Os Sertões*. Entretanto, nem ele nem Almeida apresentam uma lista completa das estruturas encontradas. Mapeamos, entre os exemplos de decassílabos e dodecassílabos apresentados por Campos, que incluem alguns exemplos de Almeida, os trechos de início e final de sentenças marcados por delimitadores (ponto, vírgula, ponto-vírgula, dois pontos, reticências e exclamação), além de sentenças completas. Resultou desse mapeamento uma base de validação formada por 35 sentenças completas, 14 trechos iniciais de sentenças e 35 trechos finais. Processamos *Os Sertões* para identificação das estruturas. Os resultados apresentaram muitas estruturas não exibidas por Almeida e por

Campos. Para comparação, usamos aquelas que estão entre os exemplos mencionados. Das 36 sentenças completas apresentadas por Campos, 35 foram identificadas pelo MIVES. A única estrutura não identificada pelo sistema foi escandida com métrica superior a 12 sílabas poéticas. A diferença resultou de variações não previstas pelo MIVES, como ecliptse, síncope e apócope. “Um primor de estatuária modelado em lama.” foi escandido por Campos como um dodecassílabo, e pelo MIVES como “Um/pri/m#or/ de es/ta/tu/#á/ria/ mo/de/l#a/do em/ l#a/ma.”,⁶ com 13 sílabas poéticas. Dois fenômenos intervocabulares foram considerados, entre as palavras “de” e “estatuária”, e entre as palavras “modelado” e “em”. Para considerar um dodecassílabo, em “estatuária” precisaríamos aplicar uma sinérese entre a 3^a e 4^a sílabas, não aplicada porque o sistema preserva a autonomia da sílaba tônica que sucede uma vogal.

Para as estruturas métricas que formam o início das sentenças, foram considerados 14 trechos iniciais de sentenças apresentados por Campos, delimitados por sinais de pontuação. Destes, 11 foram escandidas com metros equivalentes aos apresentados por Campos e três com estruturas que possuem mais de 12 sílabas. Um exemplo de estrutura não escandida, entre 10 e 12 sílabas, é a sentença “Che/ga/vam,/ es/tro/pi/a/dos./ da/ jor/na/da/ lon/ga,” classificada pelo sistema com 13 sílabas poéticas.

Almeida destacou a predominância de estruturas no fim das sentenças. De 56 estruturas, delimitadas por sinal de pontuação, 50 foram classificadas entre decassílabos e dodecassílabos; 6 não foram identificadas pelo MIVES porque suas escansões estão fora dos limites previstos, como é o caso de “exuberando floração ridente em meio da desordem tropical.”, que obteve 20 sílabas.

Se consideramos a taxa de identificação de estruturas métricas da base de validação, MIVES identificou 97,22% das sentenças completas, 78,57% no início de sentenças e 88,57% nos trechos finais. O sistema não obteve escansões entre 10 e 12 sílabas quando exigidas variações fonológicas desconsideradas (e.g., ecliptse, síncope e apócope), ou em estruturas maiores, como no exemplo mencionado de 20 sílabas.

Foram identificadas pelo MIVES, 4267 estruturas métricas em *Os Sertões*, entre decassílabos e dodecassílabos, um número muito maior de estruturas do que aquela fornecida por Campos e por Almeida. Entre decassílabos e dodecassílabos não apresentados por Campos, podemos

⁶ O símbolo # indica o posicionamento da sílaba tônica.

mencionar os heroicos “Seguiam sucessivas, incansáveis,” escandido pelo MIVES como “Se/gu#i/am/ su/ce/ss#i/vas,/ in/can/s#á/veis,” e “Calcula friamente o pugilato”, escandido como “Cal/c#u/la/ fri/a/m#en/te o/ pu/gi/l#a/to.”, e o alexandrino “Empanaram-se todas as vistas, de lágrimas...” escandido pelo MIVES como “Em/pa/n#a/ram/se/ t#o/das/as/ v#is/tas,/ de/ l#á/gri/mas...”.

Comparação entre as obras de Euclides da Cunha

Foram analisadas três obras de Euclides da Cunha – *Os Sertões*, *À Margem da História* e *Contrastes e Confrontos*. O sistema foi configurado para identificar sentenças (ou trechos de sentenças) com estruturas deca e dodecassilábicas, considerando a possibilidade de existirem diferentes padrões rítmicos para um mesmo segmento de texto. Um resumo dos resultados é apresentado na tabela 1.

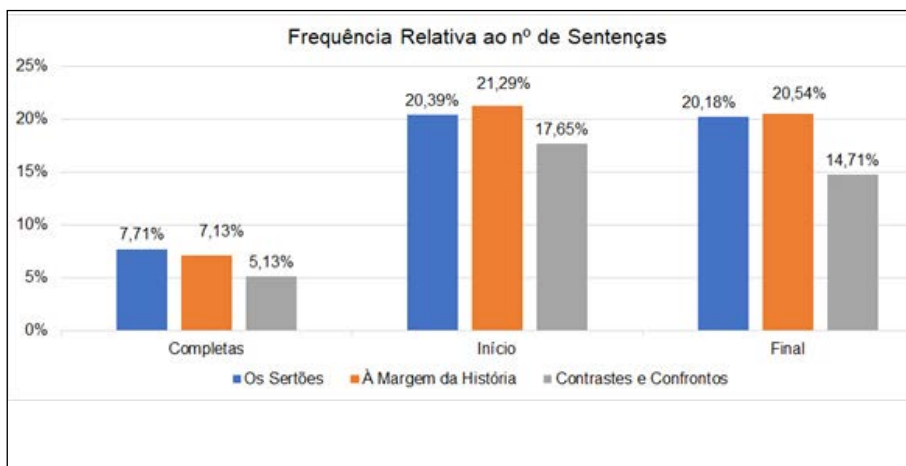
TABELA 1 – Comparação entre *Os Sertões*, *À Margem da História* e *Contrastes e Confrontos*

Parâmetros	Obra		
	<i>Os Sertões</i>	<i>À Margem da História</i>	<i>Contrastes e Confrontos</i>
Total de sentenças processadas	8564	1066	1598
Sentenças completas com estrutura métrica (total e frequência)	652 7,61%	76 7,13%	82 5,13%
Início de sentenças com estrutura métrica (total e frequência)	1746 20,39%	227 21,29%	282 17,65%
Finais de sentenças com estrutura métrica (total e frequência)	1728 20,18%	219 20,54%	235 14,71%
Quantidade de decassílabos (sentenças completas, trechos iniciais e finais)	354 (18,14%)	36 (17,14%)	43 (15,99%)
	859 (44,00%)	93 (44,29%)	134 (49,81%)
	739 (37,86%)	81 (38,57%)	92(34,20%)
	Total: 1952	Total: 210	Total: 269
Quantidade de Dodecassílabos (sentenças completas, trechos iniciais e finais)	360 (15,69%)	34 (13,28%)	40 (11,73%)
	854 (37,21%)	106 (41,40%)	156 (45,75%)
	1081 (47,10%)	116 (45,31%)	145 (42,52%)
	Total: 2295	Total: 256	Total: 341

Esta tabela exhibe uma estatística descritiva de frequências, cuja regularidade podemos chamar de “comportamento métrico-versificatório da prosa de Euclides”. Como são obras de diferentes extensões, usamos a maior delas (*Os Sertões*) como referência para comparação. Os resultados baseiam-se em frequência absoluta, para obter a quantidade total de estruturas identificadas, e frequência relativa ao total de sentenças processadas, para identificar o comportamento relacionado a estruturas de versificação por sentença processada.

Ao avaliar a densidade geral de estruturas métricas, em relação ao total de sentenças da obra (figura 4), nota-se que 48,18% das sentenças de *Os Sertões*, 48,96% das sentenças de *A Margem da História*, e 37,48% das sentenças de *Contrastes e Confrontos* apresentam estruturas, indicando alta densidade, particularmente nas duas primeiras obras. Ao comparar os valores de frequência relativa de estruturas métricas em sentenças completas, trechos iniciais e finais, nota-se que há grande proximidade entre as obras analisadas, exceção feita apenas aos casos de fins de sentença em *Contrastes e Confrontos*, de 14,71%. Mas esta obra também possui frequência inferior para sentenças completas e para início de sentenças. Ainda assim, podemos afirmar que há grande proximidade entre as obras analisadas (ver gráfico abaixo, figura 4).

FIGURA 4 – Comparativo entre três obras de Euclides da Cunha – Frequência relativa ao total de sentenças. Os rótulos “completas”, “início” e “final”, referem-se a localização das estruturas métricas no segmento analisado.



Uma importante propriedade está relacionada à grande quantidade de estruturas deca e dodecassilábicas. Os resultados observados na Tabela 1 mostram a predominância destas estruturas em relação ao total das sentenças identificadas. Elas ultrapassam 90% do total das sentenças com estruturas métricas identificadas pelo MIVES, na busca por sentenças entre 10 e 12 sílabas poéticas.

Além da estatística de frequência de estruturas métricas, podemos avaliar a distribuição de suas ocorrências ao longo da obra por uma medida de distância (quantidade de sentenças entre uma estrutura e outra), conforme vemos na tabela 2. As médias, e desvios das distâncias (quantidade de sentenças não métricas entre duas ocorrências de estruturas métricas), indicam que não existe uma regularidade na distribuição de estruturas métricas nas obras, com estruturas adjacentes (distância zero), e distantes entre si (alto valor de distância). Resultados relacionados à distância e distribuição são graficamente explorados na próxima seção.

TABELA 2 – Distâncias entre estruturas – *Os Sertões*, *À Margem da História* e *Contrastes e Confrontos*

Parâmetros	Obra		
	<i>Os Sertões</i>	<i>À Margem da História</i>	<i>Contrastes e Confrontos</i>
Distância entre sentenças (sentenças, trechos iniciais e finais): média, desvio padrão, mínimo e máximo	12,63±12,88 (0-82)	14,62±16,89(0-77)	18,38±19,61(0-109)
	4,02±4,54 (0-46)	4,33±4,37(0-27)	4,71±5,60(0-35)
	3,76±4,13 (0-30)	4,49±4,64(0-23)	5,85±6,61(0-59)

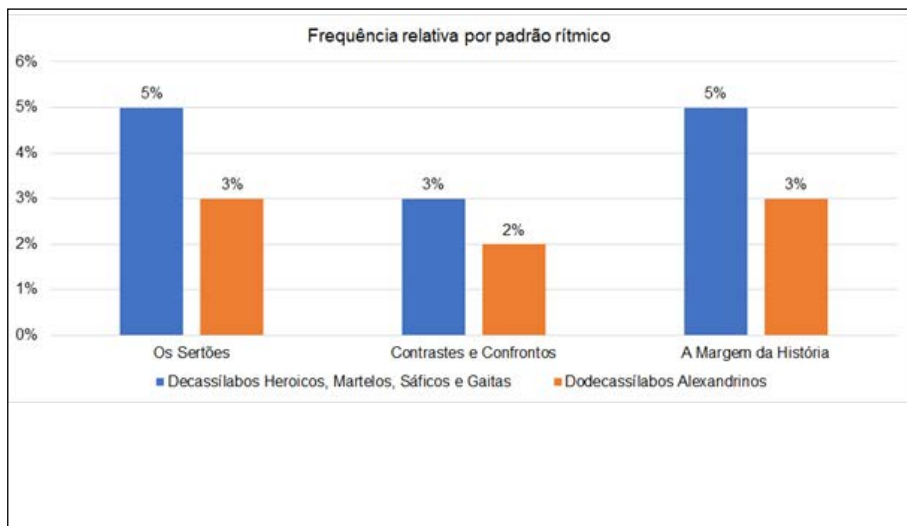
Outro resultado obtido está relacionado ao posicionamento das tônicas nas estruturas métricas. Foram identificados diversos padrões rítmicos, deca e dodecassilábicos; entre eles, heróicos/martelos, sáficos, e gaita galega (Tabela 3).

TABELA 3 – Padrões Rítmicos identificados

Obra	Heroico	Martelo	Sáfico	Gaita Galega	Alexandrino
<i>Os Sertões</i>	182	125	59	50	220
<i>Contrastes e Confrontos</i>	18	17	6	9	32
<i>À Margem da História</i>	30	14	3	4	29

Os valores absolutos de frequência dos padrões rítmicos de *Os Sertões* são consideravelmente maiores que aqueles observados nas outras obras. Mas é importante notar que as obras possuem extensões diferentes, conforme observado na Tabela 1. Em *Os Sertões* foram processadas 8564 sentenças; em *Contrastes e Confrontos*, 1598 sentenças; em *À Margem da História*, 1066 sentenças. Se analisarmos a frequência relativa dos padrões rítmicos, em relação ao número de sentenças processadas em cada obra, notamos que a frequência com que estes tipos ocorrem entre *Os Sertões* e *À Margem da História* é similar (Figura 5).

FIGURA 5 – Frequência relativa de padrões rítmicos (Decassílabos Heroicos, Martelos, Sáficos e Gaitas x Dodecassílabos Alexandrinos)



Mais resultados e potencial analítico

As informações obtidas pelo sistema incluem: identificação de estruturas, frequência absoluta e relativa de ocorrência, valores de distância entre as sentenças e valores de desvio padrão. Nesta subseção, apresentamos outros mecanismos usados para visualização e análise de resultados.

Embora a média das distâncias e dos valores de desvio padrão, como apresentados na Tabela 1, permitam avaliar uma tendência geral, e uma variação da distância entre estruturas métricas, ela não permite visualizar sua dispersão, ou distribuição, ao longo da obra; não visualizamos “regiões da obra” com maior ou menor frequência de estruturas métricas. Os gráficos de dispersão (Figuras 6, 7 e 8) permitem “situar” as estruturas métricas ao longo da obra. Nestes gráficos, o eixo central horizontal representa a sequência de sentenças na obra, e as ocorrências (sentenças com estruturas métricas) equivalem aos marcadores verticais. Áreas de maior concentração de estruturas equivalem a maior densidade de marcadores.

Os gráficos de dispersão evidenciam que as sentenças com estruturas métricas estão distribuídas ao longo de cada obra, sem notável concentração em partes específicas de cada uma. Como as obras possuem um número distinto de sentenças, cada gráfico tem uma escala diferente para o eixo horizontal. Assim, o gráfico de dispersão de *Os Sertões* aparenta ter uma maior concentração de sentenças métricas ao longo da obra, mas este é um efeito do tamanho dela, que é 8 vezes maior (quantidade de sentenças).

FIGURA 6 – Gráfico de dispersão de sentenças métricas ao longo de *Os Sertões*

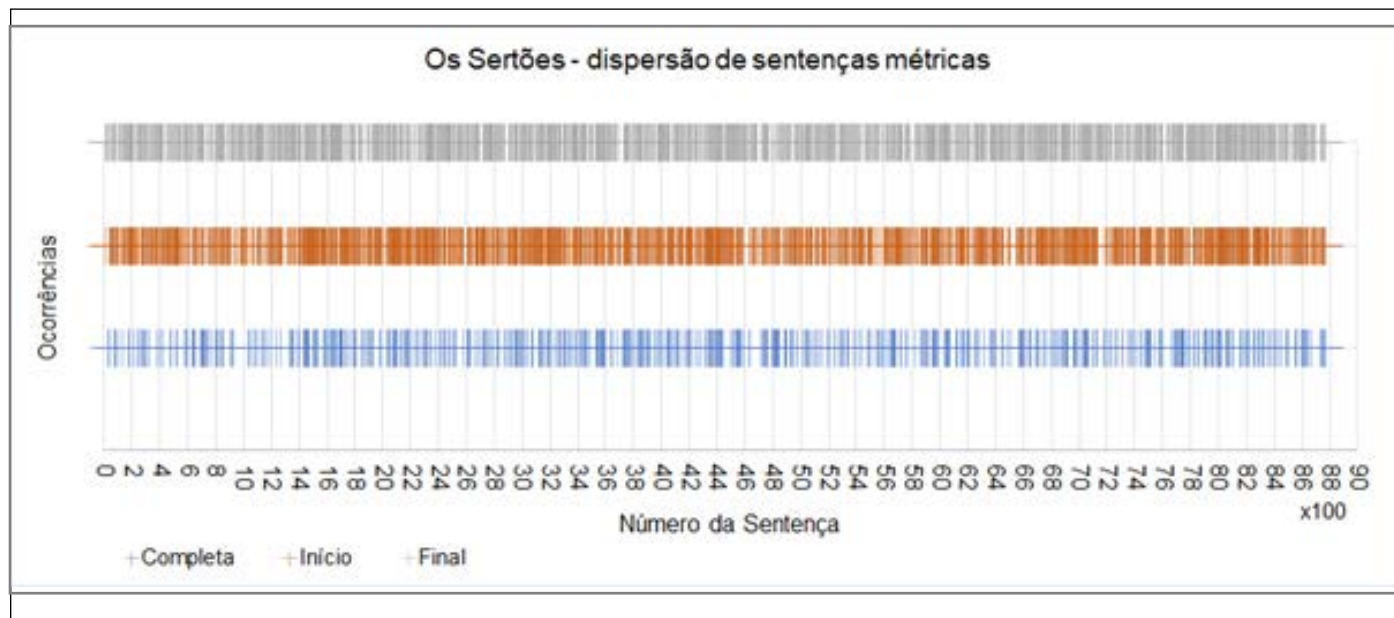
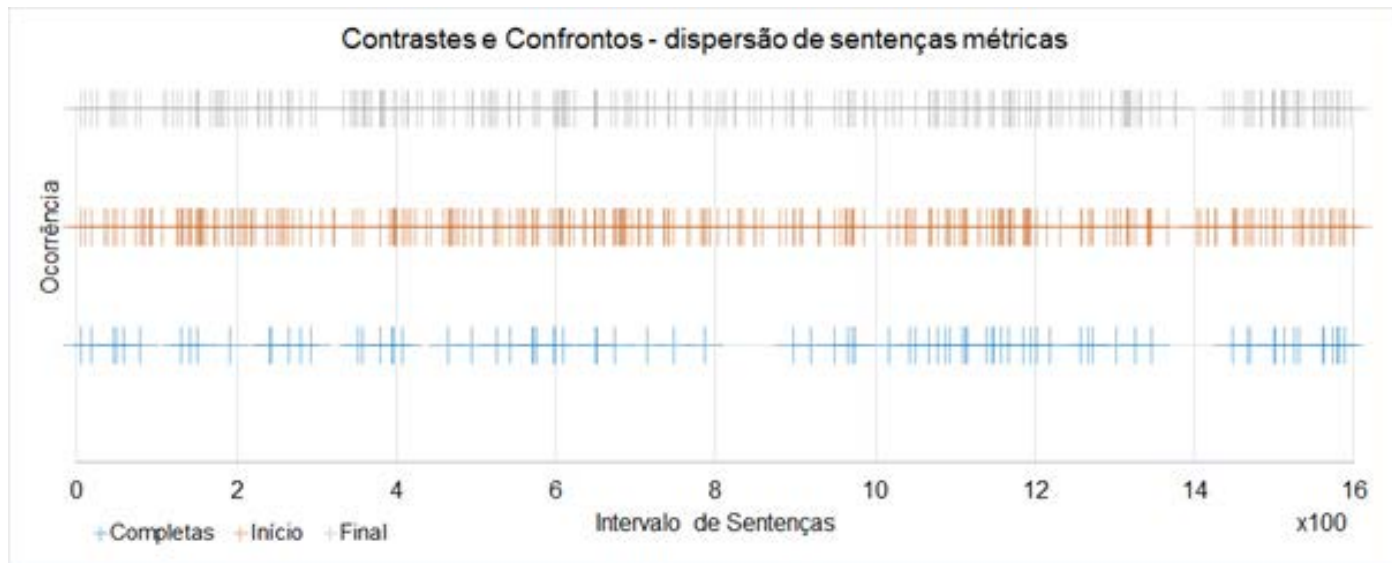


FIGURA 7 – Gráfico de dispersão de sentenças métricas ao longo de *À Margem da História*

FIGURA 8 – Gráfico de dispersão de sentenças métricas ao longo de *Contrastes e Confrontos*

Adicionalmente, com o propósito de quantificar a densidade de ocorrência de sentenças com estruturas ao longo das obras, exibimos gráficos que apresentam frequência de estruturas, em intervalos de 200 sentenças (Figuras 9, 10 e 11). Eles representam a variação contínua de frequência em intervalos regulares, exibindo regiões com maior e menor incidência de sentenças estruturadas, e sua variação ao longo da obra.

FIGURA 9 – Gráfico de distribuição de frequência absoluta de estruturas de versificação, em intervalos de 200 sentenças, para *Os Sertões*

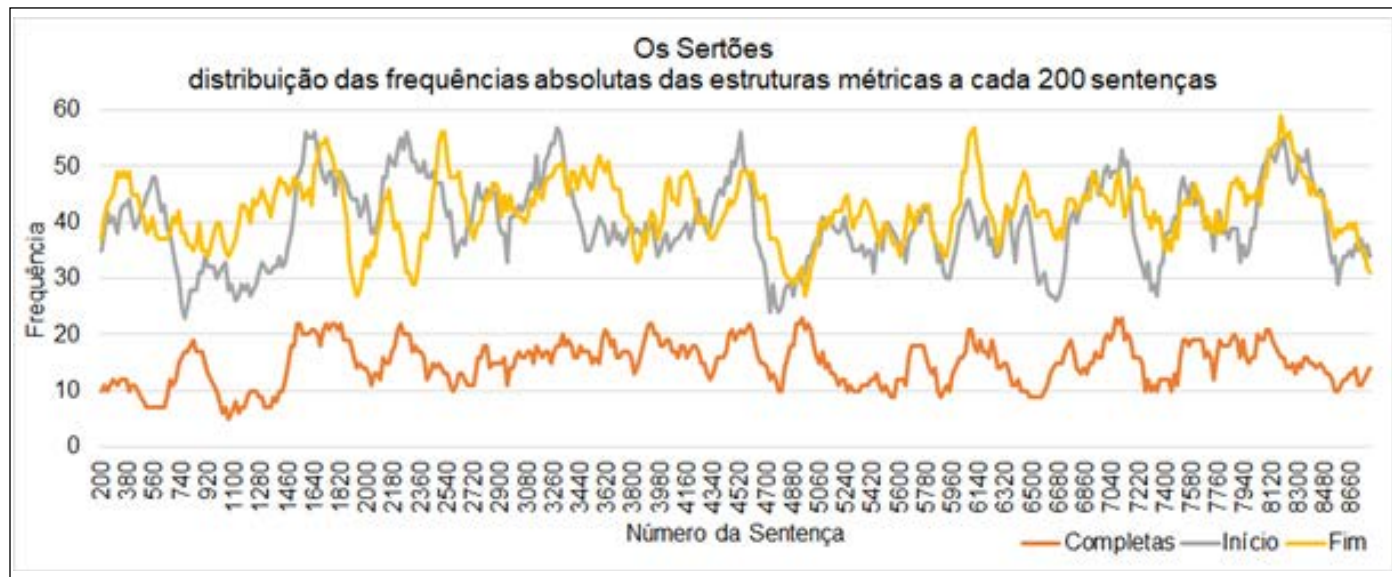


FIGURA 10 – Gráfico de distribuição de frequência absoluta de estruturas de versificação em intervalos de 200 sentenças para *À Margem da História*

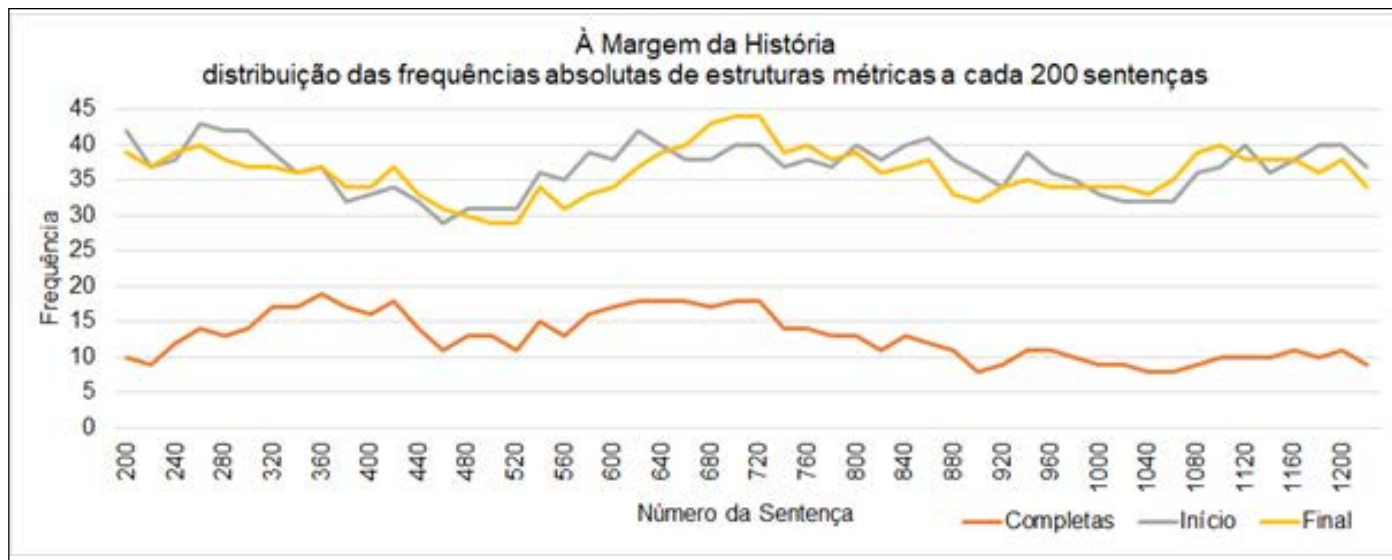
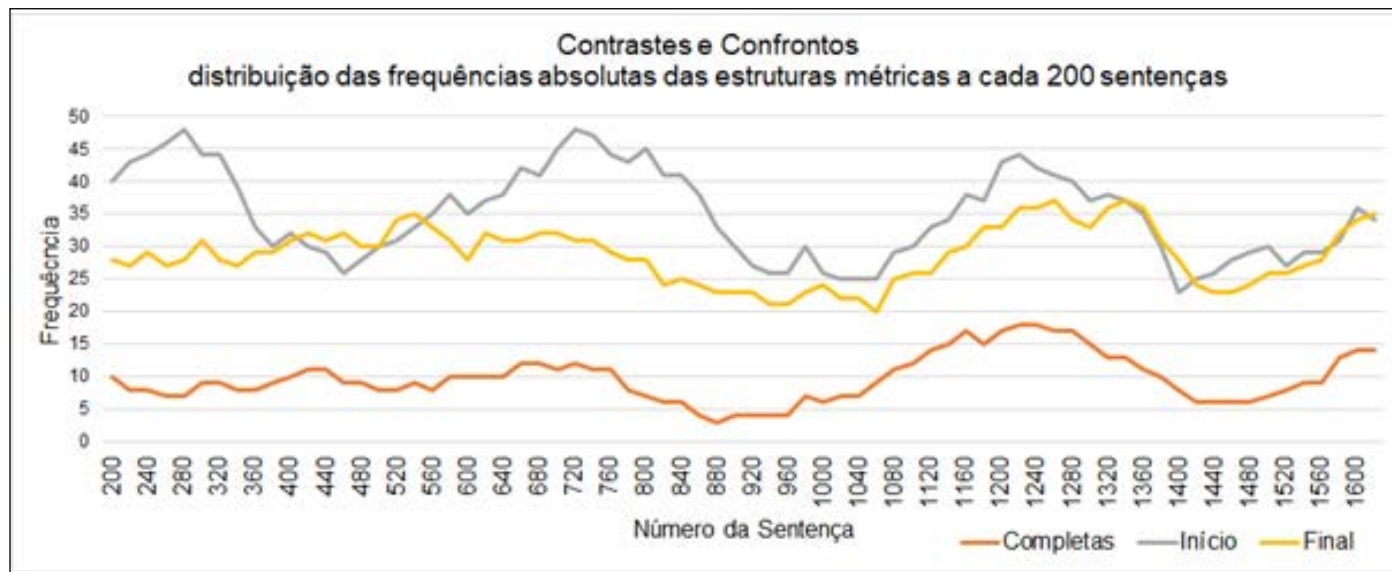


FIGURA 11 – Gráfico de distribuição de frequência absoluta de estruturas de versificação em intervalos de 200 sentenças para *Contrastes e Confrontos*



De acordo com os gráficos das Figuras 9, 10 e 11, são mais incomuns, ao longo das obras, sentenças com estruturas formadas por frases completas, quando comparadas às sentenças com estruturas métricas encontradas em inícios e finais de segmentos frásicos. Não deve ser tratado como uma coincidência o fato de que, nas três obras analisadas, a proporção de sentenças completas com estruturas métricas representam, em média, 30% do número de sentenças com estruturas métricas de início e final de sentenças. Além disso, estes gráficos indicam que a frequência de ocorrências de estruturas métricas não é uniforme. Há regiões com valores mais altos, formando picos, e outras com baixa frequência, formando vales, indicando certa alternância do fenômeno ao longo da obra. Também nota-se, nos gráficos das Figuras 7 e 8, um comportamento similar de alternância entre os valores de frequência das sentenças com estrutura métrica no início e no final. Outro aspecto notável é o comportamento similar exibido pela distribuição de frequência absoluta de estruturas entre regiões de *Os Sertões* e *À Margem da História*. Este fenômeno não tem correspondência em *Contrastes e Confrontos* (Figura 9).

4. Comentários finais

Distinto dos muitos sistemas computacionais já desenvolvidos para escandir poemas metrificados, MIVES foi concebido para escandir estruturas metrificadas na prosa, uma operação que Augusto de Campos (2010, p. 14) chamou de “leitura verso-espectral”. A automatização computacional deste processo revelou uma densidade surpreendente de estruturas de versificação em diversas obras de Euclides da Cunha. Em relação ao total de sentenças da obra, MIVES identificou estruturas em 48,18% das sentenças de *Os Sertões*, 48,96% de *À Margem da História*, e 37,48% de *Contrastes e Confrontos*, uma taxa surpreendente, quando comparada aos resultados exibidos por Augusto de Campos e Guilherme de Almeida. Podemos supor que taxas tão elevadas de estruturas possam ser encontradas na prosa literária de outros autores, outras tradições e períodos literários? Tal fenômeno ainda não foi investigado, e é o desenvolvimento “natural” deste projeto – uma análise de obras de diversos períodos para avaliar comparativamente, historicamente, o uso e o comportamento de padrões de estruturas métricas na prosa. Como afirmamos (seção 1), o que apresentamos aqui é apenas a primeira etapa de um projeto mais ambicioso.

Obviamente, MIVES pode realizar escansões em quantidade muito maior do que qualquer agente humano. Mas, ainda mais interessante, como ferramenta, é a capacidade que MIVES inaugura para identificar, quantificar e exibir padrões de distribuição de estruturas de versificação ao longo do texto, numericamente, com estatística descritiva e atributos de distância, e visualmente, através de gráficos de dispersão e frequência ao longo das obras. Não é um exagero afirmar que o sistema é capaz de abrir uma nova direção nas investigações sobre a prosa literária, em língua portuguesa.

Agradecimentos

Os autores agradecem a Mariana Salimena pela concepção e desenvolvimento da interface gráfica do sistema (MIVES).

Contribuição dos Autores

Ricardo Carvalho: concepção e desenho da pesquisa, desenvolvimento da ferramenta, obtenção de dados, análise e interpretação dos dados e redação do manuscrito. Angelo Loula: concepção, desenho e orientação da pesquisa, análise e interpretação dos dados e redação do manuscrito. João Queiroz: concepção, desenho e orientação da pesquisa, análise e interpretação dos dados e redação do manuscrito.

Referências

ALI, S. *Versificação portuguesa*. São Paulo: EDUSP, 2006.

ALMEIDA, G. A poesia d' *Os Sertões*. *Diário de São Paulo*, São Paulo, 18 ago. 1946.

ARAÚJO, P.; MAMEDE, N. Classificador de poemas. In: CONFERÊNCIA CIENTÍFICA E TECNOLÓGICA EM ENGENHARIA, 2002, Lisboa. *Actas [...]*. Lisboa: ISEL, 2002. CD-ROM

CAMÕES, L. V. *Os Lusíadas*. Biblioteca Virtual do Estudante Brasileiro / USP. Disponível em: http://www.dominiopublico.gov.br/pesquisa/ DetalheObraForm.do?select_action=&co_obra=1870. Acesso em: 30 jan. 2019.

CAMPOS, A. Transertões. In: CAMPOS, A.; ALMEIDA, G. *Poética de Os Sertões*. São Paulo: AnnaBlume, 2010.

CARVALHO, Ricardo Sena. *MIVES*: um sistema para identificação automática de padrões métricos de versificação em prosa literária brasileira. 2017. 120f. Dissertação (Mestrado em Computação Aplicada) – Universidade Estadual de Feira de Santana, Feira de Santana, 2017.

CUNHA, E. *Os Sertões*. São Paulo: Três, 1984.

CUNHA, E. *À Margem da História*. Fundação Biblioteca Nacional. Disponível em: http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2034. Acesso em: 30 jan. 2019.

CUNHA, E. *Contrastes e Confrontos*. Fundação Biblioteca Nacional. Disponível em: http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2032. Acesso em: 30 jan. 2019.

GERVÁS, P. A Logic Programming Application for the Analysis of Spanish Verse. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LOGIC, 1st., London. *Proceedings* [...]. London: Springer, 2000. p. 1330-1344. Doi: https://doi.org/10.1007/3-540-44957-4_89

JAKOBSON, R. *Dialogues*. Cambridge, MA: MIT Press, 1988.

MITTMANN, A.; WANGENHEIM, A.; SANTOS, A. L. dos. Aoidos: A System for the Automatic Scansion of Poetry Written in Portuguese. In: GELBUKH, A. (org.). *17th Computational Linguistics and Intelligent Text Processing*. Cham: Springer, 2018. p. 611-628. Doi: 10.1007/978-3-319-75487-1_46

NETO, Nelson; ROCHA, Willian; SOUSA, Gleidson. An Open-Source Rule-Based Syllabification Tool for Brazilian Portuguese. *Journal of the Brazilian Computer Society*, [S.l.], v. 21, n. 1, p. 1, 2015. Doi: 10.1186/s13173-014-0021-9

OOSTENDORP, M. Introducing a Scansion Machine for Dutch Poetry and Prose. *Loquens*, Lyon, n. 1, v. 1, e002, 2014. Doi: <https://doi.org/10.3989/loquens.2014.002>

PLAMONDON, M. R. Virtual verse analysis: analysing patterns in poetry. *Literary and Linguistic Computing*, Oxford, n. 21, v. suppl_1, p.127-141, 2006. Doi: <https://dx.doi.org/10.1093/llc/fql011>

SILVA, Denilson C.; BRAGA, Daniela; RESENDE Jr., Fernando G. V. Separação das sílabas e determinação da tonicidade no Português Brasileiro. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, XXVI., Rio de Janeiro. *Anais [...]*. Rio de Janeiro: SBrT, 2008. p. 1-5. CD-ROM.

SPINA, S. *Manual de versificação românica medieval*. São Paulo: Ateliê Editorial. 2003.