



Universidade Estadual de Feira De Santana
Programa de Pós-Graduação em Computação Aplicada

MIVES: Um Sistema para Identificação Automática de Padrões Métricos de Versificação em Prosa Literária Brasileira

Ricardo Sena Carvalho

Feira de Santana

2017



Universidade Estadual de Feira De Santana
Programa de Pós-Graduação em Computação Aplicada

Ricardo Sena Carvalho

**MIVES: Um Sistema para Identificação Automática
de Padrões Métricos de Versificação em Prosa
Literária Brasileira**

Dissertação apresentada à Universidade
Estadual de Feira de Santana como parte
dos requisitos para a obtenção do título de
Mestre em Computação Aplicada.

Orientador: Prof. Angelo Conrado Loula (UEFS)

Coorientador: Prof. João Queiroz (UFJF)

Feira de Santana

2017

Ficha Catalográfica - Biblioteca Central Julieta Carteado

C327m Carvalho, Ricardo Sena

MIVES: um sistema para identificação automática de padrões métricos de versificação em prosa literária brasileira / Ricardo Sena Carvalho. Feira de Santana, 2017.

120f.: il.

Orientador: Angelo Conrado Loula.

Coorientador: João Queiroz.

Dissertação (mestrado) - Universidade Estadual de Feira de Santana, Programa de Pós-Graduação em Computação Aplicada, 2017.

1. Versificação - Métrica - Prosa. 2. Computação - Literatura brasileira - Prosa. 3. Métrica na poesia - Computação. I. Loula, Angelo Conrado, orient. II. Queiroz, João, coorient. III. Universidade Estadual de Feira de Santana. IV. Título.

CDU: 004:869.0(81)

Está página deverá ser substituída por uma folha contendo as assinaturas dos membros da banca, e deve ser posta após a ficha catalográfica.

Abstract

In poetry, various forms of parallelism are remarkable, at various levels of description (syntactic, morphological, phonological), including metrical structures of versification. Although literary prose is not characterized by the predominance of such structures, studies indicate the presence of subliminal metric structures in the prose of Euclides da Cunha, setting the precedent for researches in a still vast field of possibilities around Brazilian literary prose. In Portuguese, the system of metrification is syllabic-accentual and the analytic decomposition of such patterns is known as scansion, and its result is not unequivocal and can vary contextually. This type of prose analysis can be defined as a non-trivial activity that, if performed in a non-automatic way, may require, depending on the size of the work, hours, days or even months of work. This work presents a system, MIVES (Mining Verse Structure), developed for automation of scansion and analysis of metric structures in Brazilian literary prose, starting with a text, extracting and processing sentences, applying possible rules and variations, and displaying results with diverse visualizations for analysis and evaluation. The system was validated through prose and poetry reference texts with known scansion, and new prose literary texts were analyzed and their metrical structures are described.

Keywords: MIVES, Versification Structure, Brazilian Literature, Scansion.

Resumo

Em poesia, são notáveis diversas formas de paralelismo, em vários níveis de descrição (sintático, morfológico, fonológico), incluindo estruturas métricas de versificação. Embora a prosa literária não seja caracterizada pela predominância de tais estruturas, estudos indicam a presença de estruturas métricas subliminares na prosa de Euclides da Cunha, abrindo precedente para pesquisas em um campo ainda vasto de possibilidades em torno prosa literária brasileira. Em português, o sistema de metrificação é silábico-acentual e a decomposição analítica de tais padrões é conhecida como “escansão”, e seu resultado não é inequívoco podendo variar contextualmente. Esse tipo de análise sobre a prosa, pode ser definida como uma atividade não-trivial que, se realizado de forma não-automática, pode exigir, dependendo do tamanho da obra, horas, dias ou até meses de trabalho. Este trabalho apresenta um sistema, MIVES (*Mining Verse Structure*), desenvolvido para automatização da escansão e análise de estruturas métricas na prosa literária brasileira, iniciando com um texto, extraíndo e processando sentenças, por possíveis regras e variações, e exibindo resultados com visualizações diversas para análise e avaliação. O sistema foi validado por meio de textos de prosa e poesia com referência de escansão conhecida, e novos textos literários em prosa foram analisados e suas estruturas métricas são descritas.

Palavras-chave: MIVES, Estrutura de Versificação, Literatura Brasileira, Escansão.

Prefácio

Esta dissertação de mestrado foi submetida a Universidade Estadual de Feira de Santana (UEFS) como requisito parcial para obtenção do grau de Mestre em Computação Aplicada.

A dissertação foi desenvolvida dentro do Programa de Pós-Graduação em Computação Aplicada (PGCA) tendo como orientador o Dr. **Angelo Conralo Loula** e co-orientador o Dr. **João Queiroz**.

Agradecimentos

A Deus, pela saúde, oportunidade e por sempre me mostrar que existe o momento certo para que tudo aconteça.

A minha família, por acreditar quando nem eu mesmo era capaz de crer que conseguiria. Em especial ao meu Pai, Nabor e minha mãe, Maria, por darem a mim o melhor que um filho espera receber dos pais. Vocês são um exemplo de luta e determinação no qual eu me inspiro todos os dias para alcançar meus objetivos. A minha esposa e companheira Carla, pela paciência, palavras de carinho e apoio incondicional nos momentos difíceis, obrigado por ser meu ponto de equilíbrio. Ao meu filho Pedro, por ser tudo para mim.

Aos meus orientadores, Angelo e João, por acreditarem, pela dedicação, cobrança e aprendizado, mas sobretudo, por me mostrarem a direção quando nem eu mesmo sabia onde deveria chegar.

Aos professores Rodrigo Calumby e Pablo Farias pelas horas dedicadas e pelos valiosos conselhos.

Aos meus amigos, vocês são parte do que hoje eu sou. Em especial a Claudia, pelo exemplo, incentivo e apoio, por tudo aquilo que fez para que eu conseguisse chegar até aqui. A Cleide, muito obrigado, seu apoio e paciência foram chaves para que agora eu tenha oportunidade de retribuir.

Aos amigos e verdadeiros companheiros que fiz durante essa jornada. Em especial, Ivonete e Thiago, sem vocês tudo teria sido mais difícil.

Obrigado a todos que contribuíram para o sucesso deste trabalho.

Sumário

Abstract	i
Resumo	ii
Prefácio	iii
Agradecimentos	iv
Sumário	v
Lista de Publicações	vi
Lista de Tabelas	vii
Lista de Figuras	x
1. Introdução	1
1.1 Objetivos	3
1.2 Organização do Trabalho.....	3
2. Fundamentação Teórica	5
2.1 Os Sertões.....	5
2.2 Os versos de Os Sertões	6
2.3 Versificação.....	7
2.4 Trabalhos Relacionados	12
3. Método	15
3.1 Etapas do MIVES.....	16
3.2 Visualização e Análise dos Resultados	25
3.3 Sistema MIVES	26
4. Resultados	33
4.1 Validação do Processo de Escansão	33
4.2 Processo de Escansão e Identificação das Estruturas de Versificação	35
4.3 Análise de Prosa de Euclides	48
4.4 Análise de Outras Obras da Prosa Literária Brasileira.....	68
5. Considerações Finais	98
6. Referências Bibliográficas	100
7. Parâmetros de escansão utilizados durante os experimentos.	102

Lista de Publicações

Carvalho, R., Loula, A., Queiroz, J. *Identificação Automática de Estruturas Métricas de Versificação na Prosa Literária de Euclides da Cunha*, STIL - Symposium in Information and Human Language Technology, 2017 (Submetido para publicação)

Carvalho, R., Loula, A., Queiroz, J. *Poetry in Prose: automatic identification of verses in brazilian literature*. In Digital Humanities 2016: Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków, pp. 756-757, 2016

Carvalho, R. S; Loula, A; Queiroz, J. *Identificação de Padrões de Versificação em Texto em Prosa da Língua Portuguesa*, WPOS - Workshop de Pós-graduação 2015, Salvador. XV Escola Regional de Computação Bahia, Alagoas e Sergipe, 2015.

Lista de Tabelas

Tabela 1: Exemplo de parágrafo antes do processo de extração de segmentos frásicos	16
Tabela 2: Exemplo de parágrafo após o processo de extração de segmentos frásicos.	17
Tabela 3: Exemplo do conjunto de parâmetros que pode ser utilizado para o processo de escansão.	20
Tabela 4: Exemplo de Conteúdo de Dicionário.....	27
Tabela 5: Relação de Estruturas de Versificação Etapa 1 - Primeiro experimento	36
Tabela 6: Os Sertões – Sentenças Completas - Ranking dos 10 Posicionamento de Tônicas	39
Tabela 7: Relação de Estruturas de Versificação Etapa 2 - Primeiro experimento	40
Tabela 8: Os Sertões – Início de Sentenças - Ranking dos 10 Padrões Rítmicos.....	43
Tabela 9: Relação de Estruturas de Versificação Etapa 3 - Primeiro experimento	44
Tabela 10: Os Sertões – Final de Sentenças - Ranking dos 10 Posicionamento de Tônicas	47
Tabela 11: Exemplo de Sentenças Completas – Contraste e Contornos.....	49
Tabela 12: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentença Completas	50
Tabela 13: Exemplo de Estruturas de Início de Sentenças – Contraste e Contornos .	51
Tabela 14: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença	53
Tabela 15: Exemplo de Final de Sentenças – Contrastes e Contornos.....	54
Tabela 16: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença.....	56
Tabela 17: Resumo Contrastes e Contornos	56
Tabela 18: Exemplo de Sentenças Completas – À Margem da História.....	57
Tabela 19: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentença Completas	59
Tabela 20: Exemplo de Início de Sentenças – À Margem da História.....	59
Tabela 21: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença	61

Tabela 22: Exemplo Final de Sentenças – À Margem da História	62
Tabela 23: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença.....	64
Tabela 24: Resumo À Margem da História.....	65
Tabela 25: Padrões rítmicos das obras analisadas de Euclides.....	66
Tabela 26: Dados Comparativos entre as Obras de Euclides	67
Tabela 27: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentença Completas.....	70
Tabela 28: Exemplo de Sentenças Completas – Canaã.....	71
Tabela 29: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença.....	73
Tabela 30: Exemplos de Estruturas de Início de Sentença - Canaã	74
Tabela 31: Exemplos de Final de Sentenças - Canaã.....	75
Tabela 32: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença.....	77
Tabela 33: Resumo das Experiências com Canaã	77
Tabela 34: Exemplo de Sentenças Completas - Recordações do Escrivão Isaías Caminha	79
Tabela 35: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentenças Completas	80
Tabela 36: Exemplo de Início de Sentenças - Recordações do Escrivão Isaías Caminha	81
Tabela 37: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentenças	83
Tabela 38: Exemplo de Estruturas de Final de Sentenças - Recordações do Escrivão Isaías Caminha	83
Tabela 39: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentenças.....	86
Tabela 40: Resumo das Experiências com Recordações do Escrivão Caminha.	86
Tabela 41: Exemplo de Sentenças Completas - Dom Casmurro	88
Tabela 42: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentenças Completas	89
Tabela 43: Exemplo de Início de Sentenças - Dom Casmurro.....	90
Tabela 44: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentenças	92
Tabela 45: Exemplo de Final de Sentenças - Dom Casmurro	93
Tabela 46: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente	

– Final de Sentenças	95
Tabela 47: Resumo das Experiências com Dom Casmurro	95

Lista de Figuras

Figura 1: Etapas Metodológicas do MIVES.....	15
Figura 2: Passos para utilização do MIVES.....	28
Figura 3: Interface para indicar a arquivo a ser processado pelo MIVES	28
Figura 4: Interface de configuração dos parâmetros de escansão do MIVES	29
Figura 5: Interface de Configuração de Metros e Tipo de Busca	29
Figura 6:Interface de Visualização e Análise dos Resultados	30
Figura 7: Interface de Detalhes do Segmento Escandido.	31
Figura 8: Gráficos de Frequência Absoluta do MIVES.....	32
Figura 9: Os Sertões - Sentenças Completas - Distribuição da Frequência Absoluta.	38
Figura 10: Os Sertões - Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto	38
Figura 11: Os Sertões - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas.....	39
Figura 12: Os Sertões – Início de Sentenças - Distribuição da Frequência Absoluta.	42
Figura 13: Os Sertões – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.	43
Figura 14: Os Sertões - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	43
Figura 15: Os Sertões – Final de Sentenças - Distribuição da Frequência Absoluta..	46
Figura 16: Os Sertões – Final de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.	46
Figura 17: Os Sertões - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.	47
Figura 18: Contrastes e Contornos - Sentenças Completas - Distribuição da Frequência Absoluta.	49
Figura 19: Contrastes e Contornos – Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto.....	49
Figura 20: Contrastes e Contornos - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas	50
Figura 21: Contrastes e Contornos – Início de Sentenas - Distribuição da Frequência	

Absoluta.....	52
Figura 22: Contrastes e Contornos – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.....	52
Figura 23: Contraste e Contornos - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	53
Figura 24: Distribuição de frequências absolutas - Contrastes e Contornos - Final de Sentenças.....	55
Figura 25: Contrastes e Contornos – Final de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.....	55
Figura 26: Contrastes e Contornos - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	56
Figura 27: A Margem da História - Sentenças Completas - Distribuição de Frequências Absolutas.....	58
Figura 28: A Margem da História – Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto.....	58
Figura 29: À Margem da História - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas.....	59
Figura 30: À Margem da História – Início de Sentenças - Distribuição das Frequências Absolutas.....	61
Figura 31: A Margem da História – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.....	61
Figura 32: À Margem da História - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	62
Figura 33: À Margem da História – Final de Sentenças - Distribuição de Frequências Absolutas.....	64
Figura 34: A Margem da História – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto.....	64
Figura 35: À Margem da História - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	64
Figura 36: Comparativo entre três Obras de Euclides - Frequência relativa ao total de sentenças.....	65
Figura 37: Comparativo dos Padrões Rítmicos sobre Prosa de Euclides da Cunha...	66
Figura 38: Canaã – Sentenças Completas - Distribuição da Frequência Absoluta....	69
Figura 39: Canaã – Sentenças Completas - Distribuição das sentenças Métricas ao longo do texto.....	70
Figura 40: Canaã - Sentenças Completas- Frequência Absoluta - Posicionamento das Tônicas.....	70
Figura 41: Canaã – Início de Sentenças - Distribuição de Frequências Absolutas....	72
Figura 42: Canaã – Início de Sentenças - Distribuição das sentenças métricas ao longo	

do texto	72
Figura 43: Canaã - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas	73
Figura 44: Canaã – Final de Sentença - Distribuição da Frequências Absoluta.....	76
Figura 45: Canaã – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto.....	76
Figura 46: Canaã - Final de Sentenças- Frequência Absoluta - Posicionamento das Tônicas	77
Figura 47: Recordações do Escrivão Caminha – Sentenças Completas - Distribuição da Frequência Absoluta.....	80
Figura 48: Recordações do Escrivão Caminha – Sentenças Completas - Distribuição das sentenças métricas ao longo do texto.....	80
Figura 49: Recordações do Escrivão Caminha – Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas.....	81
Figura 50: Recordações do Escrivão Caminha – Início de Sentenças - Distribuição de Frequências Absolutas.....	82
Figura 51: Recordações do Escrivão Caminha – Início de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto.....	82
Figura 52: Recordações do Escrivão Caminha – Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	83
Figura 53: Recordações do Escrivão Caminha – Final de Sentença - Distribuição de Frequência Absoluta.....	85
Figura 54: Recordações do Escrivão Caminha – Final de Sentenças - Distribuição das sentenças métricas ao longo do texto.....	85
Figura 55: Recordações do Escrivão Caminha – Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	86
Figura 56: Dom Casmurro – Sentenças Completas - Distribuição da Frequências Absolutas.....	89
Figura 57: Dom Casmurro – Sentenças Completas - Distribuição das sentenças métricas ao longo do texto.....	89
Figura 58: Dom Casmurro – Sentenças Completas- Frequência Absoluta - Posicionamento das Tônicas	90
Figura 59: Distribuição de frequências absolutas – Dom Casmurro – Início de Sentenças	91
Figura 60: Dom Casmurro – Início de Sentenças - Distribuição das sentenças métricas ao longo do texto.....	92
Figura 61: Dom Casmurro – Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas	92
Figura 62: Dom Casmurro – Final de Sentença - Distribuição da Frequências Absolutas.....	

.....	94
Figura 63: Dom Casmurro – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto	94
Figura 64: Dom Casmurro – Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.....	95

Capítulo 1

Introdução

Dentre expressões da escrita, destaca-se a arte literária ou literatura. Neste campo, duas grandes categorias se distinguem, em termos classificatórios – prosa e poesia. Sua distinção, nem sempre de clara demarcação, baseia-se na identificação de uma grande variedade de estruturas e processos.

Diferente da poesia, a prosa não precisa obedecer a rimas ou ritmos regulares [Silva 2015, p.25]. A poesia é lida, mais acentuada ou explicitamente, com para-morfismos fonológicos, como os paralelismos sonoros (rimas, trocadilhos e paronomásias), métricos e rítmicos, morfossintáticos e gramaticais [Bernucci e Hardman 2009]. Na poesia, “o ritmo segue uma modulação mais atenta às unidades melódicas - semântico - emotivas que a sintaxe, e muitas vezes os segmentos frásicos articulam-se com relativa simetria, que lembra a regularidade do verso” [Moisés, 2002,p. 356]. Em poesia, são notáveis diversas formas de paralelismo, em vários níveis de descrição (sintático, morfológico, fonológico), incluindo estruturas métricas de versificação. Em português, o sistema de versificação é silábico-acental, ou seja, conta-se o número de sílabas dos versos, em seguida, localizam-se as sílabas fortes, tônicas [Goldstein 2006, p. 27]. A cadência e a alternância entre sílabas fortes e fracas estabelecem um certo número de padrões que, combinados as repetições posicionais das sílabas, cria segmentos internos, que seguem regras de versificação ou metrificação. A forma do verso é determinada pela combinação das sílabas, acentos e pausas, contando-se até a última sílaba acentuada do verso [Cunha e Cintra 2008, p. 686]. A decomposição analítica de tais padrões é conhecida como “escansão”, e seu resultado não é inequívoco podendo variar contextualmente. Fazer a escansão ou escandir um verso consiste em decompô-lo em sílabas métricas [Goldstein 2006, p. 106] ou, como também designadas pela literatura, sílabas poéticas. A sílaba poética pode ser diferenciada em forma da sílaba gramatical como está descrito em nosso referencial teórico.

Entre a prosa de linguagem denotativa e a poesia de linguagem conotativa, temos a prosa literária. “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” [Jakobson 1985, p.106]. Em textos, literários, no qual as palavras nem sempre são usadas em seu sentido literal, elementos da comunicação usados na prosa misturaram-se a instrumentos ou recursos fortemente encontrados na poesia, como rimas, trocadilhos e aliterações. Embora a prosa literária não seja caracterizada pela predominância de estruturas métricas, estudos indicam a presença

de estruturas métricas subliminares na prosa de Euclides da Cunha. Em 1946, Guilherme de Almeida publicou um estudo inédito sobre Os Sertões, um texto escrito em prosa, em que revelou aspectos poéticos (métricos e fonológicos), sobre livro Os Sertões, de Euclides da Cunha, publicado em 1902. Quase meio século depois, precisamente em 1996, Augusto de Campos publicou um artigo intitulado Transertões, mais uma vez destacando o estilo de escrita em versos que marca inúmeros trechos do livro de Euclides. Em seus trabalhos, Almeida seguido por Campos, revelam a presença de estruturas heterométricas (de diferentes metros) de versificação, padrões característicos do português. Os decassílabos e alexandrinos estão entre os metros mais usados em combinações e posições variadas. A variedade de padrões encontrados, desprezada “a metrificação estrita” e admitida “maior liberdade rítmica” [Campos, 2010, p. 29], cria surpreendentes zonas de tensão, “áreas pregnantas de poesia em trechos significativos de sua prosa” [Campos, 2010]. Um problema científico pode prescindir facilmente da contingência da descoberta de [Almeida, 1946] e [Campos, 1996] e portanto, podem existir versos em outras obras de Euclides da Cunha ou sob outras obras da prosa literária brasileira.

O trabalho realizado por Almeida e Campos, podem se concretizar em um processo de investigação em que a extensão da obra e a experiência do pesquisador são determinantes para o tempo necessário para se fazer esse tipo de pesquisa, se feitas manualmente podem implicar em dias, ou até mesmo meses de trabalho, até ser concluída, dentre outros, por exigir do pesquisador, além de uma leitura atenta, um estudo pormenorizado de possíveis métricas sobre centenas de trechos da obra literária em análise. Por outro lado, hoje a utilização do computador, com a aplicação de técnicas específicas para cada objetivo, pode permitir que buscas e análises, antes feitas manualmente e, em muitos casos, durante um longo período de tempo, possam ser realizadas de forma mais rápida por meio da automatização de algumas etapas, reduzindo, desta forma, o esforço e o tempo necessário para a realização da tarefa.

Diversas pesquisas e diferentes técnicas ao longo de décadas foram desenvolvidas com o objetivo de criar mecanismos computacionais para busca, processamento e análise de textos. Com a ampliação do universo de documentos digitais disponíveis, a diversidade e qualidade dos algoritmos, e a velocidade com que os computadores passaram a desempenhar tarefas tais como armazenamento, manipulação, transferência e recuperação da informação, trazendo inúmeros benefícios e atendendo os mais variados tipos de domínio (eg. acadêmicos industriais ou científicos), a exemplo do trabalho desenvolvido por [Willians et al 2014] que tem como objetivo acelerar as análises de papíros gregos antigos.

Dentro do universo de mecanismos e aplicações voltados ao processamento de textos, diversos sistemas desenvolvidos recentemente em linguística computacional concentraram-se na automatização da escansão de poemas em versos metrificados, a exemplo de [Mittmann 2016]. Nestes casos, o principal propósito é a escansão de estruturas métricas previamente classificadas como verso. Neste trabalho, “Paralelismo” é uma ideia-chave para caracterizar nosso problema. A “estrutura da poesia é um paralelismo contínuo” [Jakobson 1985]. Os paralelismos são correspondências contínuas em vários níveis de descrição, tais como: i) morfológicos; ii) sintáticos e; iii) fonológicos. O fenômeno não aparece na prosa. Mas a prosa literária é um fenômeno de especial interesse porque é um fenômeno intermediário [Jakobson 1985]. Na prosa examinada

aqui, paralelismos de estruturas rítmicas variadas (decassilábicas, dodecassilábicas, etc.) são “reveladas” através de procedimentos computacionais.

Em nossos levantamentos, sobretudo para o português, não foram encontrados trabalhos da mesma natureza que esta dissertação se refere. Mesmo para a poesia, em que as estruturas de versificação são imprescindíveis, existe uma escassez de trabalhos relacionados, com exceção ao trabalho realizado por [Mittmann 2016], como pode ser constatado em nosso referencial teórico.

A pesquisa que esta dissertação apresenta torna possível inúmeras possibilidades de descobertas no extenso acervo literário que existe em português, sobretudo na literatura portuguesa, já que, “o Brasil é um país de escritores. Grandes, médios e pequenos. Homens e principalmente mulheres, numa proporção cada vez maior. Aí se inclui o gênero poesia, com adeptos numerosos, como indicam os admiradores de Castro Alves, Cecília Meirelles, Carlos Drummond de Andrade e Vinícius de Moraes, entre outros” [Niskier 2003]. A possibilidade de análise da mesma natureza feita por Augusto de Campos e Guilherme de Almeida em *Os Sertões*, com o auxílio do computador, em outras obras, poderá reduzir custo ou esforço de estudiosos da área de linguística ou da literatura para trazer à luz, estruturas heterométricas de versificação em obras escritas em prosa literária brasileira.

1.1 Objetivos

Geral

Propor uma metodologia computacional para identificação automática de estruturas de versificação da prosa literária brasileira.

Específicos:

- Construir uma ferramenta computacional para implementar a metodologia proposta;
- Conceber e implementar uma interface de análise para os resultados obtidos;
- Validar o processo de escansão com textos de métrica já conhecida;
- Aplicar o método a outras obras escritas por Euclides da Cunha;
- Realizar a busca por estruturas subliminares de versificação em obras literárias do mesmo período e comparar os resultados obtidos;
- Disponibilizar para comunidade acadêmica a ferramenta computacional desenvolvida.

1.2 Organização do Trabalho

Os próximos capítulos deste documento estão organizados da seguinte forma:

O Capítulo 2 , *Fundamentação Teórica*, apresenta os conceitos elementares para compreender a solução desenvolvida bem como os trabalhos relacionados.

No Capítulo 3 , *Metodologia*, apresentamos as estratégias utilizadas e caminhos

percorridos na construção do MIVES.

No Capítulo 4 , Resultados, apresenta os resultados, tanto no processo de mineração em busca das estruturas de verificação, como naqueles em torno das possibilidades de análise entregue pelo MIVES ou usuário.

No Capítulo 5 , apresentamos as considerações finais e indicamos possíveis desdobramentos deste estudo.

Capítulo 2

Fundamentação Teórica

Sobre a automatização do processo de identificação de estruturas subliminares de versificação em texto em prosa da literatura portuguesa, por ser uma área de pesquisa ainda inexplorada, não foram encontrados trabalhos científicos ou soluções baseadas em *software* para o português. A referência para esta pesquisa reside nos trabalhos realizados por dois poetas brasileiros, [Almeida 1946] e [Campos 1996]. Retomando ao que já foi descrito no Capítulo 1, os trabalhos revelam estruturas de versificação de diferentes metros presentes em *Os Sertões – Campanha de Canudos*, livro escrito por Euclides Pimenta da Cunha.

A descrição de *Os Sertões*, bem como dos trabalhos realizados por Guilherme de Almeida e Augusto de Campos estão presentes, respectivamente nas seções 2.1 e 2.2. O restante deste capítulo está organizado da seguinte forma: na seção 2.3 são apresentados os conceitos de versificação características da poesia. Já na seção 2.3.1 são apresentadas as classificações métricas do verso conforme sistema de versificação portuguesa, muito empregado no Brasil antes da eclosão dos movimentos modernistas no Brasil e em Portugal. Por fim, na seção 2.4 são listados e brevemente descritos, os trabalhos relacionados a aplicação de recursos computacionais no processamento de estruturas textuais relacionadas com a escansão de segmentos frásicos caracterizados como verso.

2.1 Os Sertões

Publicada em 1902, *Os Sertões*, dentre muitas classificações atribuídas, recebeu o título de “Bíblia da nacionalidade” [Abrel 2009], como obra-prima da prosa, escrita em língua portuguesa, e já observara José Veríssimo,

...um vasto e profundo tratado de ciência humanas, inovador nos métodos, nas teorias e nas conclusões, nas pesquisas de campo e na reformulação de conceitos, sobre os estudos do homem, do meio e da sociedade e cultura brasileiros, especialmente dos sertanejos [Brandão 2002, p.11].

Resultado da observação *in loco* daquela inexplicável e sanguinolenta guerra civil, no fundo dos sertões ignotos da Bahia, a obra, dentre os inúmeros adjetivos atribuídos, também é considerada como uma narrativa histórica do confronto armado ocorrido entre os anos de 1896 e 1897 entre os moradores de um arraial e o Exército Brasileiro. O primeiro, movido pela necessidade de sobrevivência diante das dificuldades impostas

pela própria natureza e pela alta e intolerante cobrança de impostos, e o segundo movido, dentre outras, pela pressão de latifundiários locais e pela necessidade de afirmação de uma república recém promulgada que via no arraial liderado por Antônio Vicente Mendes Maciel, o Antônio Conselheiro, uma ameaça aos interesses da jovem república.

Dividido em três partes, A Terra, O Homem e A Luta, o livro é uma referência e fonte de pesquisa histórica deste confronto. Um texto jornalístico, já que Euclides na época fora enviado ao local do confronto como correspondente do Jornal o Estado de São Paulo. Quatro décadas depois de publicado, estudiosos trouxeram a luz características do texto que revelaram estruturas de versificação das mais variadas. “Os Sertões é um desses casos em que o brilho precoce e estrondoso de um livro acabaria por ofuscar o restante da produção, seja em prosa, seja em poesia” [Bernucci 2009, p. 13].

2.2 Os versos de Os Sertões

Em 1946, o poeta brasileiro Guilherme de Almeida publicou um estudo [Almeida, 1946] a respeito do poeta Euclides e dos padrões estruturados de versos identificados por ele na prosa de Os Sertões. De acordo com o trabalho de [Almeida, 1946], existe na prosa Euclidiana versos de todos os gêneros poéticos, dentre eles: o heroico, o lírico, o descritivo, o bucólico, o satírico e o epigramático, concebidos através de estruturas de versificação de vários padrões rítmicos, aparentemente e com maior frequência nos finais de parágrafos. São exemplos destes versos os decassílabos: “num ondear longínquo de chapadas”; “os recessos das matas opulentas”, ou ainda o alexandrino, “o antagonismo permanente das montanhas”. [Almeida 1946] ainda apresenta versos de diferentes metros localizados ou destacados em diferentes pontos do texto, como em: “O sertanejo é, antes de tudo, um forte”, um dodecassílabo em frase destacada que dá início ao capítulo III do livro de Euclides.

A análise de Almeida e a discussão em seu artigo, intitulado de A poesia de Os Sertões, é apresentada em torno dos aspectos poéticos de Os Sertões. Não são apresentados resultados quantitativos, em números de estruturas que encontrou, mas qualitativos, já que, descreve elementos da poesia presentes no livro de Euclides.

Cinquenta anos depois, em 1996, foi publicado um trabalho por Augusto de Campos [Campos 1996], ampliando o que já fora anunciado por Guilherme de Almeida [Almeida 1946] e revelando várias outras estruturas de versificação na prosa Euclidiana. Decassílabos e Dodecassílabos estão entre os padrões métricos encontrados com maior frequência, em várias combinações e posições. A diversidade de padrões identificados, desconsiderando “a metrificação estrita” e admitindo “mais liberdade de ritmos” [Campos 1996], criam, de acordo com Campos, surpreendentemente zonas de tensão, “áreas gestantes de poesia em trechos significativos de prosa.” [Campos 1996]. A exemplo da musicalidade imposta pela aliteração presente neste fragmento de texto destacado por [Campos 1996, p. 3]: “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;”.

O resultado do artigo sobre esta obra de Euclides é a descrição e a apresentação de algumas estruturas de texto ou segmentos frásicos encontrados no livro, conceituadas

por [Campos 1996] como “*os aspectos poéticos da linguagem Euclidiana*”. São apresentados cerca de 185 das mais 700 estruturas de versificação, entre decassílabos e dodecassílabos que [Campos 1996] afirma ter encontrado no livro.

procurando privilegiar aquelas situações frásicas em que a trilha métrica soa conatural ao ritmo da fala, cheguei a mais de 500 decassílabos significativos, com predominância dos sáficos (acentuados na 4| e 8| sílaba) e heroicos (acentuados na 6|) e pouco mais de duas centenas de dodecassílabos (dentre os quais muitos alexandrinos perfeitos) [Campos 1996, p. 1].

As estruturas de verificação apresentadas por Campos [1996] podem ser encontradas em inúmeros pontos da obra, no início da frase, mas em sua maioria como terminais de parágrafos, e como já afirmará [Campos 1996], “Os Sertões começa com um decassílabo heroico: “O planalto central do Brasil desce...” e como já observara Guilherme de Almeida, termina com um solene alexandrino “...as linhas essenciais do crime e da loucura””. Além das estruturas que marcam o início e os finais de alguns parágrafos, Campos também elenca inúmeros estruturas de versificação concebidas por Euclides, como parágrafos autônomos, como em “O sertanejo é, antes de tudo, um forte”.

Além desses locais de maior concentração de versos, [Campos 1996] apresenta outras formações em frases extensas, em que são reveladas sequências como em “O planalto central do Brasil desce nos litorais do Sul, em escarpas inteiriças, altas e abruptas”, uma frase que se inicia com um decassílabo heroico e termina como um dodecassílabo.

Além dos aspectos já apresentados, o artigo escrito por [Campos 1996] revela outros aspectos que aproximam a prosa de Euclides em Os Sertões às palavras poesia e poema, como exemplo, a existência de outros segmentos frásicos que ultrapassam a contagem ou a métrica tradicional, “seria fácil demonstrar em Euclides análoga independência do número métrico, já que o movimento rítmico de suas frases ultrapassa em muito as contagens convencionais”, [Campos 1996]. [Almeida 1946] também deixa claro em seu artigo ter encontrados outros elementos na obra de Euclides, e não é o metro, como ritmo e musicalidade, que aproxima a prosa do escritor da poesia.

2.3 Versificação

Para Spina [2003], o verso é a linha do poema, que pode possuir diversas extensões. No português, o sistema de versificação é silábico acentual. Conta-se o número de sílabas de cada verso e a alternância entre sílabas fortes (acentuadas) e fracas [Spina 2003]. As alternâncias estabelecem um certo número de arranjos que combinados em suas partes, ou em repetições posicionais de sílabas acentuadas e fracas, cria segmentos internos, estabelecendo padrões de versificação ou metrificação. A esta sucessão de sílabas fortes e fracas, em intervalos regulares, chamamos de ritmo.

A análise do verso inclui, entre outras atividades, a contagem de suas sílabas poéticas. Versos normativamente identificados recebem uma classificação (as classificações métricas comuns do verso são apresentadas na seção 2.4) de acordo com o número de sílabas encontradas. A esse processo atribui-se a conceito de escansão, e é através dele que certas características do verso são identificadas, incluindo sua estrutura métrica. A forma metrificada do verso é determinada pela combinação de sílabas, acentos e pausas, contadas até a última acentuada [Cunha e Cintra 2008]. Segundo Cunha e Cintra [2008]:

Este sistema de escandir o verso até o último acento tônico, segundo o modelo francês, foi introduzido em Portugal no século XVIII por Miguel de Couto Guerreiro (Tratado de versificação portuguesa. Lisboa, Francisco Luiz Ameno, 1784, p.6-7) mas se vulgarizou em nossa métrica pelo enorme prestígio de Antônia Feliciano de Castilho, que o acolheu no seu Tratado de metrificação portuguesa. Lisboa, Imprensa Nacional, 1851. Anteriormente contava-se os versos portugueses, à semelhança dos espanhóis e italianos, com base no verso grave, ou seja, considerando sempre no cômputo a existência de uma sílaba átona depois da última tônica. [Cunha e Cintra 2008, p. 686]

Durante o processo de escansão são identificados os fenômenos fonológicos (hiato intervocabular, sinalefa, elisão e crase são exemplos), destacadas, entre outras coisas, pela utilização de células fonêmicas ou pela alternância entre sílabas fortes e fracas. Estes fenômenos fonológicos podem impactar na medida da palavra (número de sílabas da palavra) e, conseqüentemente, na medida do verso. Essa propriedade é importante para destacar que o número de sílabas poéticas, ao final da escansão, pode diferir do número de sílabas gramaticais sobre o mesmo segmento de texto.

Tomemos o verso de Olavo Bilac como exemplo: “E a alma de sonhos povoada eu tinha...”. Esse verso, segundo [Cunha e Cintra, 2008], após a escansão, possui dez sílabas, já que fora separado como segue: “E a al/ma/ de/ so/nhos/ po/vo/a/da eu/ ti/nha...”. Pode-se observar neste exemplo, um encurtamento no total do número de sílabas. Ao comparar o processo de escansão com a contagem silábica gramatical, aplicado ao mesmo verso de Bilac, obtemos um total de quatorze sílabas. Esse fenômeno não é usual. Há casos em que o número de sílabas encontradas é igual à quantidade de sílabas obtidas a partir da separação silábica gramatical, como pode-se observar neste outro verso de Olavo Bilac: “Cheguei. Chegaste. Vinhas fatigada”, escandido por [Cunha e Cintra, 2008, p. 687], na forma: “Che/guei./ Che/gas/te. / Vi/nhas/ fa/ti/ga/da”. Segundo ele, “se abandonarmos a última sílaba átona, de qualquer forma que o leiamos haverá dez sílabas até a última tônica”.

A quantidade de sílabas poéticas de um segmento de texto, contabilizada ao final de um processo de escansão, resulta de fenômenos fonológicos, principalmente, como: Diérese, Sinérese, Sinalefa, Elisão e Crase. A análise, no processo de escansão, da estrutura de formação interna das palavras, e a organização destas ao estabelecerem ligações rítmicas, permite observar a ação conjunta destes fenômenos fonológicos. Segundo Cunha e Cintra [2008]:

- Sinalefa: A primeira vogal torna-se uma semivogal, que passa a formar ditongo com a seguinte. Exemplificando: Por definição gramatical “fatigado eu” possui cinco sílabas - fa-ti-ga-do eu. Já no verso, pela necessidade de imposição rítmica, em alguns casos, “o” de fatigado pode tornar-se uma semivogal e formar um ditongo com o “e” de eu. Neste caso, a separação silábica resultaria em quatro sílabas poéticas.
- Elisão: A primeira vogal pode desaparecer na pronúncia diante de uma vogal de natureza diversa. Para exemplificar, na pronúncia “fatigada eu”, mais uma vez o ritmo, imposto pela leitura desta pronúncia em um verso, pode resultar na supressão do som da vogal “a” tornado mais aparente o som emitido pelo “e” de eu, resultando, na escansão, em quatro sílabas ao invés de cinco.
- Crase: A primeira vogal pode ser igual à segunda e com ela fundir-se. A crase acontece pela fusão de sons entre vogais contíguas iguais. É o que ocorre em “Tinhas a alma”, quando o “a” desaparece da pronúncia, reduzindo em uma unidade a

quantidade de sílabas.

Sinalefa, Elisão e Crase definem encontros vocálicos que acontecem entre palavras, no mesmo verso. Estes encontros, de sílabas em palavras distintas, que “normalmente” resultariam em duas sílabas, podem sofrer um efeito de ditongação por imposição do ritmo do verso. Além destas, outras propriedades precisam ser observadas, pois explicam o alongamento ou redução da quantidade de sílabas de uma palavra, nos casos de Sinérese, Dírese e Crase. O primeiro, sinérese, ocorre pela necessidade de transformar um hiato em um ditongo, enquanto o segundo, a dírese, acontece pela transformação de um ditongo em hiato. Na Crase, observamos a fusão de duas vogais idênticas, em sílabas gramaticais distintas.

No verso, “Mas o seu olhar magoado”, que é heptassílabo de Augusto Gil, segundo [Cunha e Cintra, 2008], assume a forma escandida: “Mas/ o/ seu/ o/lhar/ ma/goa/do”. Neste verso, a palavra “magoado”, gramaticalmente separada como “ma-go-a-da”, passou, na segunda e terceira sílabas por um efeito de ditongação, deixando de ser uma palavra tetrassilábica, e se tornando um trissílabo (ma-go-a-do), caracterizando, desta forma, devido a exigência métrica, um caso de Sinérese.

A exigência métrica pode fazer, por exemplo, da palavra “saudade”, trissílabo (sau-da-de) se tornar um tetrassílabo (sa-u-da-de). É o que pode ser observado nos seguintes versos, segundo [Cunha e Cintra, 2008], heptassílabos:

1. Eu não quero, nem brincando,
2. Dizer adeus a ninguém:
3. Quem parte leva saudades,
4. Quem fica saudades tem.
5. A ausência tem uma filha,
6. Que se chama saudade:
7. Eu sustento mãe e filha,
8. Bem contra minha vontade.

A fim de preservar a uniformidade no número de sílabas, ou o isossilabismo, nos dois primeiros versos (linhas 3 e 4), onde aparece a palavra saudade, esta se manteve como um trissílabo. Porém, se a mesma regra for aplicada a palavra saudade no verso: “Que se chama saudade” (linha 6), este se tornará um hexassílabo. Portanto, neste último verso, por exigência métrica, a palavra “saudade” foi tratada como um tetrassílabo, sua primeira sílaba, deixou de ser um ditongo (sau) para hiato (sa-u). O processo de alongamento silábico sofrido pela palavra “saudade” é definido como Dírese.

Semelhante a sinérese, a crase reduz o número de sílabas de uma palavra. Vejamos um exemplo neste verso decassílabo de Castro Alves: “Quando eu passo no Saara amortalhada”. Aqui, a palavra Saara teve os dois “aa” contíguos unidos, resultando em um decassílabo (Quan/do eu/ pas/so/ no/ Saa/ra a/mor/ta/lha/da).

Na Crase, vogais que gramaticalmente ficariam em sílabas distintas, têm suas sílabas unidas em uma só, por serem idênticas e se apresentarem na mesma palavra de forma contígua. É o que pode ser visto neste decassílabo de Castro Alves: “Quando eu passo no Saara amortalhada”. A palavra Saara que gramaticalmente seria um trissílabo (sa-a-ra) teve os dois “aa” contíguos unidos (saa-ra) resultando, ao final da escansão, em um verso decassílabo (Quan/do eu/ pas/so/ no/ Saa/ra a/mor/ta/lha/da).

Além da Sinérese, Diérese e Crase, que são fenômenos de redução ou ampliação do número sílabas de uma palavra, [Cunha e Cintra, 2008] descreve a Aférese, Síncope e Apócope como fenômenos presentes em versos. Segundo Cunha [2008]:

- Aférese: Supressão de sons no início da palavra. Como no verso de Castro Alves “‘Sta/mos/ em ple/no mar. . . / Do/ fir/ma/men/to”. Neste caso, foi empregada a forma ‘Stamos por Estamos.
- Síncope: Supressão de sons no meio da palavra. Como acontece em “Es/p’ran/ças/al/tas. . . / Ei-/las/ já/ tão/ ra/sas”. Neste verso, de Casimiro de Abreu, a pronúncia Esp’ranças por Esperanças exemplifica a ocorrência da síncope em um verso.
- Apócope: Supressão de sons no fim da palavra. Em “Ar/tis/ta-/ cor/ta o/már/mor/de/ Car/ra/ra”, a forma “mamor” substitui “mármore”.

Todos as questões relacionadas com a escansão, abordadas nos parágrafos acima, permitem, entre outras coisas, uma atribuição de classes relacionadas ao número de sílabas encontradas e ao posicionamento das tônicas no verso. Essas classificações são explicadas na subseção imediatamente a seguir.

2.3.1 Classificação Métrica dos Versos

Durante a escansão são identificados os posicionamentos das tônicas, permitindo classificar o verso em subcategorias, como será explicado na listagem a seguir. Os termos que tipificam o verso são compostos gregos formados por sua forma numeral seguido do elemento sílabo. A seguir, uma breve descrição dos tipos mais comuns, apresentados por [Cunha e Cintra 2008]:

- Monossílabo: Versos de uma sílaba
- Dissílabo: Verso de duas sílabas
- Trissílabo: Verso de três sílabas
- Tetrassílabo: Verso de quatro sílabas, pode apresentar três variações, com ocorrência de sílabas tônicas nas posições:
 - 2ª e na 4ª (mais comum) sílaba;
 - 1ª e na 4ª sílaba;
 - 4ª sílaba;
- Pentassílabo: Verso de cinco sílabas, pode apresentar quatro variações, com ocorrência de sílabas tônicas nas posições:
 - na 2ª e na 5ª (mais comum) sílaba
 - na 1ª, na 3ª e na 5ª sílaba
 - na 3ª e na 5ª sílaba
 - na 1ª e na 5ª sílaba
- Hexassílabo: Verso de seis sílabas, pode apresentar seis variações, com ocorrência de sílabas tônicas nas posições:
 - na 2ª, na 4ª, e na 6ª sílaba

- o na 2ª e na 6ª sílaba
- o na 4ª e na 6ª sílaba
- o na 1ª, na 4ª e na 6ª sílaba
- o na 1ª, na 3ª e na 6ª sílaba
- o na 3ª e na 6ª sílaba
- Heptassílabo: verso de sete sílabas, pode apresentar oito variações, com ocorrência de sílabas tônicas nas posições:
 - o na 1ª, na 3ª, na 5ª e na 7ª sílaba
 - o na 3ª, na 5ª e na 7ª sílaba
 - o na 1ª, na 3ª e na 7ª sílaba
 - o na 3ª e na 7ª sílaba
 - o na 4ª e na 7ª sílaba
 - o na 2ª, 4ª e na 7ª sílaba
 - o na 1ª, 2ª, 4ª e na 7ª sílaba
 - o na 2ª, na 5ª, e na 7ª sílaba.
- Octossílabo: verso de oito sílabas, podem ser apresentados em dez variações, com ocorrência de sílabas tônicas nas posições:
 - o na 2ª, na 4ª, na 6ª e na 8ª sílaba;
 - o na 2ª, na 4ª e na 8ª sílaba;
 - o na 4ª e na 8ª sílabas, mas podendo ter ou não a 1ª sílaba acentuada;
 - o Variante também do primeiro tipo, com acentuação interna apenas na 4ª sílaba, ou na 1ª e na 4ª.
 - o na 2ª, na 6ª e na 8ª sílaba;
 - o na 1ª, na 3ª, na 5ª e na 8ª sílabas;
 - o na 3ª, na 5ª e na 8ª sílabas;
 - o na 1ª, na 5ª e na 8ª sílabas;
 - o na 2ª, na 5ª e na 8ª sílaba
 - o na 3ª, na 6ª e na 8ª, podendo ter a 1ª sílaba também forte.
- Eneassílabo: O verso de nove sílabas está classificado em dois tipos:
 - o Eneassílabo Anapéstico, que apresenta acentuação na 3ª, na 6ª e na 9ª sílaba;
 - o O Eneassílabo com acento interno fundamental na 4ª sílaba, que, por exigência idiomática, recebe forçosamente um outro na 6ª ou na 7ª sílaba.
- Decassílabo: Verso de dez sílabas com duas subclassificações e com predominância de tônicas conforme apresentado a seguir:

- o O decassílabo heroico: tônicas na 6ª e na 10ª sílaba, podendo existir ocorrências de sílabas tônicas secundárias como na 8ª sílaba e em uma das quatro primeiras sílabas;
- o O decassílabo chamado sáfico, que apresenta acentuação na 4ª, na 8ª e 10ª sílaba, podendo, naturalmente, ter a 1ª ou a 2ª também fortes;
- Hendecassílabo: foi muito usado pelos nossos poetas românticos numa variação no posicionamento das tônicas sempre uniforme, ou seja, com acentuação na 2ª, na 5ª, na 8ª e na 11ª sílaba.
- Dodecassílabo: verso de doze sílabas é mais conhecido por verso Alexandrino. Também chamado de clássico possuindo sílabas tônicas nas posições:
 - o 6ª e na 12ª sílaba;

Existe também o Alexandrino Clássico – o chamado Tetrâmetro -, que apresenta, além dos acentos principais (na 6ª e na 12ª) dois outros secundários (na 3ª e na 9ª).

Conforme descrito no início desta subseção, o objetivo é de apresentar as estruturas de versificação utilizadas na língua Portuguesa. As descrições dos metros foram retirados, em muitos casos, na íntegra da obra de [Cunha e Cintra 2008]. Estes metros foram destacados pela sua importância estrutural ser uma das bases da identificação das estruturas de versificação em textos em prosa.

2.4 Trabalhos Relacionados

Conforme descrito no início desta seção, embora não tenham sido encontrados trabalhos relacionados diretamente com esta pesquisa, foram realizados trabalhos que executam a escansão automatizada de versos por meio de técnicas computacionais. Os trabalhos são apresentados sequencialmente pela ordem de relevância ou proximidade com esta pesquisa e estão agrupados de acordo com os objetivos apresentados, análise de poemas, ferramentas de auxílio a composição de poemas e estudos que envolvem a geração automática de poesia.

Mittmann [2016], apresenta o Aaios, um sistema capaz de realizar a escansão automática de poemas escritos em português. Segundo o autor, o sistema é baseado em regras, o Aaios recebe um poema escrito na língua portuguesa e executa a escansão automaticamente. Além do poema escandido, o Aaios fornece uma análise de rimas do poema. O Aaios é composto por vários módulos. Inicialmente o sistema recebe como entrada um arquivo XML do tipo TEI (*Text Encoding Initiative*). Cada verso nesse arquivo é extraído e submetido ao pré-processamento, que envolve a normalização do texto, toda a pontuação é removida e todas as letras são convertidas em letras minúsculas. Na etapa seguinte, o sistema produz uma ampla transcrição fonética que é produzida para cada palavra em cada verso. As palavras então, são processadas por um módulo de prosódia que aplica regras para os encontros entre vogais que representam sinalefas, elisões e crases. Além disso, o módulo de prosódia é empregado para realizar uma escansão com base em prosódia natural. O módulo de ritmo analisa o resultado produzido pelo módulo, utilizando prosódias nativas para todos os versos do poema e descobre a métrica do poema bem como padrões de sílabas tônicas. Para finalizar, o módulo de rima analisa a transcrição fonética final de todos os versos em

um poema e atribui-lhes etiqueta que correspondem ao esquema de rima. Para o autor o sistema apresentou excelentes resultados (acima 98% de precisão) nos experimentos a que submeteram o sistema.

Araújo e Mamede [2002] propuseram um protótipo para um classificador de poemas para o português europeu. A proposta dos autores é baseada em uma arquitetura capaz de auxiliar estudantes e escritores, na classificação de poemas com base nos conceitos de estrofe, verso, sílaba e rima. Segundo os autores, a ferramenta poderá funcionar como uma ferramenta didática de apoio ao estudo de poesia nas escolas e para poetas que poderão ter tarefas de versificação dos poemas simplificadas. A arquitetura agrega um processo central chamado de Processo Classificador de Poemas, este é responsável por coordenar quatro módulos que compõem o protótipo construído. Um módulo que realiza a interface com a aplicação externa designa-se por módulo de interface externa. A aplicação externa gera transcrições fonéticas e divisões silábicas e o seu funcionamento está baseado no sistema DIXI, este tem como base um sistema de regras para realizar transcrição fonética e a divisão silábica. O sistema conta com o léxico que inclui nas transcrições fonéticas e nas divisões silábicas informações adicionais para facilitar o processo de classificação. Outro módulo o MIC – módulo identificador de conceitos, identifica os conceitos de poesia como a rima do poema. A classificação final do poema é realizada pelo módulo identificador de regras. Neste artigo, o autor não apresenta resultados quantitativos de suas experiências apenas exemplos que ilustram o resultado do processamento de um poema. Nas conclusões o autor afirma que os resultados foram alcançados mediante a classificação de diferentes poemas.

Gervas [2000] apresenta uma solução baseada em *software* para automatizar o processo de análises de corpus poéticos formados por sonetos escritos em espanhol. As regras implementadas pelo algoritmo são fornecidas para capturar regras formais que governam a poesia formal em espanhol. Basicamente, a proposta de Gervas constitui como uma ferramenta para estudo da poesia literária espanhol, como um instrumento pedagógico para estudantes fornecendo capacidades de análises sobre textos poéticos. A metodologia utilizada no trabalho é formada por passos que envolvem: a) a decomposição de versos em sílabas gramaticais; b) estas sílabas são reorganizadas em sílabas poéticas, é feita a localização da sílaba tônica, uma contagem da sílaba métrica considerando os casos de sinalefa, uma análise formal da poesia envolvendo a redistribuição do acento prosódico de acordo com padrão rítmicos precisos e por último; c) extração e identificação do tipo rima. A proposta de Gervas foi testada em um conjunto de 64 sonetos clássicos. Os sonetos foram escolhidos por serem poemas de forma estrutural rígida, todos possuem 11 sílabas. Segundo Gervas, o sistema mostrou bons resultados em uma amostra que classificou como pequena.

A relação que as pesquisas apresentadas neste capítulo possuem com o trabalho descrito nesta dissertação reside na capacidade dos algoritmos em processar estruturas textuais utilizando regras de escansão. Como será visto no Capítulo 3 , Método, o MIVES também realiza o processamento de estruturas textuais utilizando regras de escansão. Porém, diferente dos trabalhos apresentados que possuem a unidade do poema previamente definida, para realizar a escansão, o MIVES precisa lidar com outras variáveis, tais como: as diferentes variações de um texto em prosa, a falta de um contexto para determinar os metros, onde começa ou termina o verso ou estrutura de versificação, com o intuito de apresentar ao seu usuário estruturas de versificação

aparentemente veladas, ou pouco aparentes, em textos da prosa literária brasileira.

Capítulo 3

Método

Como ferramenta computacional de prospecção inicial sobre os estudos relacionados a aspectos versificatórios subliminares da prosa literária brasileira, foi desenvolvido o sistema MIVES (*Mining Verse Structure*). MIVES incorpora regras e definições levantadas durante esta pesquisa a respeito de estratégias de escansão de versos para o português. As definições estabelecidas para o processo de busca por estruturas de versificação em textos em prosa, são baseadas nos tratados de versificação para o português do Brasil (ver [Spina 2003]), respeitando as regras e definições amplamente difundidas em livros do gênero e gramáticas.

A automatização do processo de identificação de estruturas de verificação em prosa é realizada por um sistema que executa etapas bem definidas. A Figura 1 apresenta uma visão geral das etapas de processamento realizadas pelo MIVES.

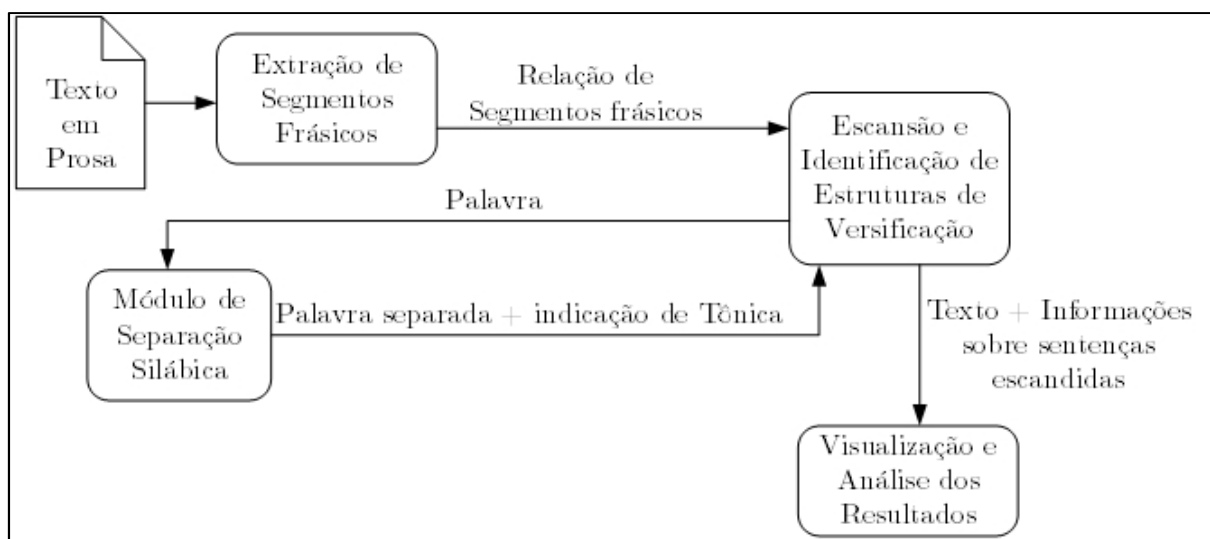


Figura 1: Etapas Metodológicas do MIVES

Tudo começa com a entrada do texto em prosa. Em seguida o sistema divide o texto em segmentos frásicos. Estes são submetidos ao processo de escansão e identificação de estruturas de versificação. Durante a escansão são feitas consultas para obter o resultado da separação gramatical, módulo de separação silábica responsável por fazer a separação gramatical prévia e a indicação da sílaba tônica de cada palavra. Ao final

os resultados destas etapas são enviados para a visualização e análise dos resultados.

As próximas seções deste capítulo apresentam um detalhamento das etapas de processamento executadas pelo MIVES durante o processamento do texto. Em seguida uma descrição dos recursos disponibilizados ao usuário para análise e visualização dos resultados. Por fim, apresenta uma descrição a respeito da arquitetura do sistema.

3.1 Etapas do MIVES

As principais etapas de processamento apresentadas nesta seção correspondem aos passos executados pelo MIVES, incluindo a abordagem utilizada para apresentação dos resultados. A Figura 1 apresenta uma visão geral da sequência de atividades executadas pelo MIVES. Todo processo se dá mediante as necessidades de busca declaradas pelo usuário, como: os metros mínimos e máximos que deseja buscar; onde as buscas devem ser concentrar (e.g. na sentença completa, no início ou apenas no fim das sentenças); ou ainda se preferir, informar quais encontros entre vogais devem ser considerados como fenômenos fonológicos intra e Intervocabulares. Além destas, o usuário deve informar a obra literária em formato digital TXT. Nas subseções a seguir estão descritas como, e em que ordem, colaboram e são executadas cada uma das etapas de processamento executadas pelo MIVES.

3.1.1. Extração de Segmentos Frásicos

A primeira etapa executada pelo *software* é a extração de segmentos frásicos. Entende-se por segmento frásico uma sequência de palavras (letras e números) adjacentes que terminam com um ponto final de sentença. Este ponto pode ser um ponto de segmento, ponto parágrafo ou um ponto final, que funciona como um delimitador para um segmento frásico. Pontos que acompanham algarismos romanos ou pronomes de tratamento, por exemplo, não são considerados como ponto final de sentença. Nesta etapa, outros sinais de pontuação não são considerados como delimitadores.

Como exemplo, é apresentado na Tabela 1 o fragmento extraído do arquivo TXT do livro, Os Sertões:

Tabela 1: Exemplo de parágrafo antes do processo de extração de segmentos frásicos

“O seu aspecto recorda, vagamente, à primeira vista, o de guerreiro antigo exausto da refrega. As vestes são uma armadura. Envolto no gibão de couro curtido, de bode ou de vaqueta; apertado no colete também de couro; calçando as perneiras, de couro curtido ainda, muito justas, cosidas às pernas e subindo até as virilhas, articuladas em joelheiras de sola; e resguardados os pés e as mãos pelas luvas e guarda-pés de pele de veado — é como a forma grosseira de um campeador medieval desgarrado em nosso tempo. Esta armadura, porém, de um vermelho pardo, como se fosse de bronze flexível, não tem cintilações, não rebrilha ferida pelo sol. É fosca e poenta. Envolve ao combatente de uma batalha sem vitórias... A sela da montaria, feita por ele mesmo, imita o lombilho rio-grandense, mas é mais curta e cavada, sem os apetrechos luxuosos daquele. São acessórios uma manta de pele de bode, um couro resistente, cobrindo as ancas do animal, peitorais que lhe resguardam o peito, e as joelheiras apresilhadas às juntas. ”

Após a extração, o sistema armazena os segmentos frásicos organizados sequencialmente (na mesma ordem em que aparecem no texto), separados por linha e numerados conforme o exemplo de saída apresentado na Tabela 2.

Tabela 2: Exemplo de parágrafo após o processo de extração de segmentos frásicos.

Número	Sentença
1	O seu aspecto recorda, vagamente, à primeira vista, o de guerreiro antigo exausto da refrega.
2	As vestes são uma armadura.
3	Envolto no gibão de couro curtido, de bode ou de vaqueta; apertado no colete também de couro; calçando as perneiras, de couro curtido ainda, muito justas, cosidas às pernas e subindo até as virilhas, articuladas em joelheiras de sola; e resguardados os pés e as mãos pelas luvas e guardapés de pele de veado — é como a forma grosseira de um campeador medieval desgarrado em nosso tempo.
4	Esta armadura, porém, de um vermelho pardo, como se fosse de bronze flexível, não tem cintilações, não rebrilha ferida pelo sol.
5	É fosca e poenta.
6	Envolve ao combatente de uma batalha sem vitórias... A sela da montaria, feita por ele mesmo, imita o lombilho rio-grandense, mas é mais curta e cavada, sem os apetrechos luxuosos daquele.
7	São acessórios uma manta de pele de bode, um couro resistente, cobrindo as ancas do animal, peitorais que lhe resguardam o peito, e as joelheiras apresilhadas às juntas.

Como um dos objetivos deste trabalho é a identificação de segmentos frásicos, estes podem ter ocorrências em frases completas, início ou término de frases, isolar o segmento frásico define a unidade de dados para o processamento posterior.

3.1.2. Separação Silábica

Um dos processos importantes para a escansão é a separação silábica e identificação das tônicas. Para [Cunha e Cintra 2008, p. 690], “relativamente à contagem das sílabas no interior das palavras temos de considerar, em primeiro plano, os fatores de ordem gramatical”. A sílaba gramatical se diferencia da sílaba poética pelo fato deste último ter sua forma influenciada por características que emergem do processo de escansão. Essas características influenciam na medida da palavra (sinérese, diérese e crase) e na

relação das sílabas entre as palavras do mesmo verso (ver subseção 2.3). No MIVES, o processo de escansão, determina quais palavras devem ser submetidas a separação silábica gramatical realizada pelos artefatos apresentados nesta seção.

A separação silábica das palavras acontece de uma forma individual, por palavra. Todo o método é controlado pelo processo de Escansão e Identificação das estruturas de Versificação, descrito na seção seguinte, que envia palavras individuais para separação silábica gramatical. Essa separação silábica inicial é realizada pelo o algoritmo desenvolvido por [Neto 2015] que funciona tomando como base 20 regras fonológicas propostas inicialmente por Silva et. al, [2008]. A ideia principal implementada no algoritmo é que todas as sílabas possuem uma vogal como núcleo, que pode ser cercada por consoantes ou por outras vogais (semivogais ou glide). Segundo Neto [2015], o algoritmo foi testado com um extrato de 10.000 palavras escolhidas aleatoriamente na base de dados do CETENFolha. Neto [2015], diz ainda que a eficiência deste algoritmo no processo de separação obteve rendimento médio aproximado de 99.14%. Além do processo de separação silábica, este algoritmo encontra a vogal tônica da palavra.

Devido ao processamento intensivo de regras e questões de integração de tecnologias (como será descrito na seção 3.3 O Sistema Mives), o processo de separação silábica gramatical é lento. Por esse motivo, incorporamos uma estrutura de mapa baseada em um algoritmo de *hash* onde, as palavras processadas pelo algoritmo de Neto [2015] são armazenadas como índices e sua correspondente separação silábica com a marcação de tônica como valor. Um exemplo de como as palavras são organizadas neste dicionário pode ser visto na Tabela 4. Essa estrutura de *hash* oferece o benefício de não reprocessar palavras presentes neste dicionário entregando o valor solicitado de forma mais rápida se comparada a ter que reprocessá-la a todo instante que for necessário. Além disso, essa estrutura contribui para resolver inconsistências no processo de separação silábica e identificação de tônicas, já que o próprio autor afirma existirem fragilidades que precisam ser melhor estudadas, quando afirma que “As falhas encontradas no algoritmo proposto se concentraram nos encontros vocálicos em final de palavra, como em “euforia” (eu-fo-ria), cujo o ditongo crescente “ia” é mantido erroneamente na mesma sílaba.”. No que se refere à indicação de tônicas, o algoritmo também apresentou algumas inconsistências. Em nossas experiências, as palavras “roubarem” e “sàbiamente” tiveram suas tônicas indicadas na primeira sílaba, quando ambas deveriam receber a indicação de tonicidade na penúltima sílaba, já que são paroxítonas. Além de garantir mais desempenho no processo de separação silábica, o dicionário permite, através de intervenção manual, que essas anomalias sejam corrigidas pelo próprio usuário.

Nesta etapa, uma sentença como, “É o homem permanentemente fatigado” torna-se “É /o h#o /mem/ per/ma/nen/te/m#en/te/ fa/ti/g#a/do. ”, no qual / indica um separador de sílaba e # é um marcador de sílaba tônica. Este processo é controlado pelo módulo de Escansão e Identificação de Estruturas de Versificação, descrito na subseção a seguir.

3.1.3. Escansão e Identificação das Estruturas de Versificação

Conforme descrito da seção 2.3, ascansão é o processo de separação, marcação e contagem de sílabas poéticas. É através deste processo que se identificam as “estruturas ou padrões de versificação” e conseqüentemente o número de sílabas poéticas, o

posicionamento das tônicas e a classificação do verso. O método descrito aqui representa o núcleo do MIVES. Os fenômenos a serem considerados, que influenciarão na identificação e classificação de estruturas de versificação são definidas pelo usuário. No processo de escansão, as preferências do usuário influenciam diretamente no resultado final. Não se trata de fazer com que o sistema incorpore todas as características levadas em consideração pelo usuário durante um processo manual de escansão, mas o de permitir que usuário parametrize o que considera relevante em situações inerentes ao processo de escansão. Trata-se de uma tentativa de aproximar as decisões que serão tomadas pelo *software* do que é esperado pelo usuário.

Sinérese, diérese, elisões e outras, são fenômenos fonológicos que emergem da escansão. Estes fenômenos são observados nos casos em que há “encontros”, em sequências, de vogais e/ou semivogais. Nestes casos, a separação silábica gramatical pode ser alterada por restrições métricas, gerando um alongamento ou uma diminuição no número de sílabas das palavras. Situações semelhantes podem acontecer também entre palavras. Duas ou mais sílabas adjacentes pertencentes a palavras distintas podem ser consideradas apenas como uma, este fenômeno impõem uma diminuição no número de sílabas nas estruturas de versificação. Por esse motivo, é necessário “instruir” o sistema a respeito do que deve ser feito quando da ocorrência de determinados encontros entre vogais.

Os fenômenos que influenciam no processo de separação de sílabas durante a escansão não ocorrem de forma rígida. Ou seja, podem ou não ser aplicadas durante a escansão, gerando variações alternativas para o resultado deste processo, inclusive com metros diferentes. Com objetivo de atender a essas questões, o MIVES permite que o usuário configure os parâmetros de escansão indicando, por fenômeno fonológico e para cada par de vogais adjacentes, apresentados na Tabela 3, o qual deverá ser o comportamento do sistema. Para isso, o usuário pode indicar para um determinado par de vogais adjacentes uma das três opções a seguir:

- Tente: O sistema pode aplicar ou não o fenômeno ao qual está relacionado, tentando manter-se dentro das restrições métricas;
- Sempre: Independente do efeito causado, o sistema sempre aplicará o tipo de fenômeno ao qual está relacionado;
- Nunca: O sistema nunca deverá aplicar aquele tipo de fenômeno caso este esteja presente no texto.

O MIVES não foi concebido para realizar o processamento de estruturas textuais utilizadas na poesia, baseada em versos alocados em estrofes, assim o contexto de outros versos pode influenciar a escansão para obter uniformidade. Por esse motivo, consideramos que o processo de escansão no MIVES deve obedecer às restrições métricas da faixa de metros informada pelo usuário, que define valores mínimos e máximos do metro. Os casos de diminuição ou aumento do número de sílabas poéticas, serão aplicados sempre com o objetivo de obter uma escansão dentro do intervalo de metros que o usuário deseja, avaliando variações e combinações possíveis. Assim, o sistema busca obter escansões para todos os valores de metro dentro da faixa de valores de metro (mínimo e máximo), informadas pelo o usuário.

Para que um determinado encontro vocálico, ou vogais adjacentes no mesmo segmento, possa ser tratado pelo sistema como um possível fenômeno fonológico, este precisa fazer

parte da relação de parâmetros de escansão utilizada pelo sistema. A Tabela 3, apresenta um exemplo de como estes parâmetros são agrupados pelo MIVES, a coluna Sequência lista exemplos de pares de vogais que podem ser considerados durante a escansão. Estes são aplicados pelo sistema mediante à necessidade de adequação métrica das estruturas que estão em processamento.

Tabela 3: Exemplo do conjunto de parâmetros que pode ser utilizado para o processo de escansão.

Fenômeno Fonológico	Sequências
Diérese	e-u, a-u, e-i, u-i, a-i, o-i, i-u, o-í
Sinérese	a-e, i-o, e-o, i-u, u-a, a-o, u-e, o-a, o-e, i-a, u-o, e-a, i-e
Sinalefa, Elisão e Crase	u-a, u-e, o-a, u-i, o-e, i-a, o-i, u-o, o-é, e-a, i-e, a-a, e-e, i-i, o-o, u-u, a-e, e-i, i-o, a-i, e-o, i-u, a-o, e-u, a-u

O Algoritmo 1, descreve as etapas realizadas pelo MIVES durante o processo de escansão para cada valor de metro do intervalo especificado (min a máx.). As entradas recebidas pelo algoritmo são o segmento a ser escandido (S) e o número de sílabas poéticas que se deseja encontrar (M). O segmento frásico é inicialmente dividido em palavras. A partir da primeira, e para cada palavra, é efetuada uma consulta no dicionário para obter a palavra separada gramaticalmente, caso não esteja presente no dicionário, é acionado o algoritmo de separação silábica que devolve a palavra separada gramaticalmente. Na ordem, são localizadas as vogais adjacentes Intravocabulares e aplicados nestes encontros de vogais, os casos de diérese (linha 5, Algoritmo 1), crase, seguido pelas situações de sinérese (linha 6, Algoritmo 1), obtendo-se a separação silábica poética da palavra. Cada palavra, representada no Algoritmo 1 pela variável p, é tratada como um objeto com atributos, como o número de sílabas, número de diéreses e número de sinéreses. Quando os métodos aplicarDierese, e aplicarSinereese, por exemplo, são executados, o estado dos atributos de cada palavra é alterado e as sílabas da palavra são redefinidas, assumindo uma forma poética, de acordo com o fenômeno fonológico identificado. A cada palavra escandida, é feita a contagem da sílaba poética até a palavra processada no momento. As palavras escandidas vão compondo o segmento métrico escandido (linha 9 ou 16, Algoritmo 1). A medida que as palavras são incorporadas ao segmento, são localizadas também as vogais adjacentes Intervocabulares e verificadas os fenômenos fonológicos (elisões, sinalefas e crases) (linha 13, Algoritmo 1) e atualizadas a contagem de sílabas poéticas.

Algoritmo 1: Escansão

Entrada: Segmento frásico S, metro a ser considerado M

Saída: objeto verso escandido

```

1: i ← 0
2: palavras[] ← quebrarSegmentoEmPalavras(S)
3: para todo p ∈ palavras faça
4:     dicionario(p)
5:     aplicarDierese(p)
6:     aplicarSinereese(p)

```

```

7:     se i = 0 então
8:         qtdS ← p.getNumeroDeSilabas()
9:         verso.add(p)
10:        pAnte ← p
11:        i ← 1
12:    else
13:        eexisteLigacao(pAntes, p) então
14:            qtdS ← qtdS - 1
15:        qtdS ← qtdS + p.getNumeroDeSilabas()
16:        verso.add(p)
17:        pAnte ← p
18:    fimse
19:    se qtdS = M então
20:        retorne verso
21:    senão se qtdS > M então
22:        break
23:    fimse
24: fimPara
25: se qtdS > M e (qtdS - verso.getNumeroDiereses() ≤ M) então
26:     para todo p ∈ verso faça
27:         se p.existeDierese() então
28:             repita
29:                 d.ignoraDierese()
30:                 qtdS ← qtdS - 1
31:                 ate p.existeDierese() e qtdS ≠ M
32:         se qtdS = M então
33:             retorne verso
34:         fimse
35:     fimPara
36: fimse
37: se qtdS < M e (qtdS + verso.getNumeroDeSinereses() + verso.getNumeroDeElisoes() ≥ M)
38: então
39:     para todo p ∈ verso faça
40:         se p.existeSinerese() então
41:             repita
42:                 p.ignoraSinerese()
43:                 qtdS ← qtdS + 1
44:                 ate p.existeSinerese() e qtdS ≠ M
45:         se qtdS = M então
46:             retorne verso
47:         fimse
48:     fimPara
49: enquanto verso.existeElisao() e qtdS < M faça
50:     verso.ignoraElisao()
51:     qtdS ← qtdS + 1
52: fimEnquanto
53: retorne verso
54: fimSe
55: retorne verso

```

Pelo fato do MIVES escandir segmentos frásicos de tamanhos ou quantidade de palavras desconhecidos, ao final do processamento de cada palavra, é verificado se a métrica até então atingida pela escansão satisfaz os valores de metro, número de sílabas

poéticas que se deseja encontrar (M) (linha 19, Algoritmo 1).

Conforme descrito na seção 2.3, Versificação, é comum no processo de escansão a necessidade de adequação métrica do verso. Durante o processo de escansão realizado pelo MIVES, caso a métrica alcançada não esteja no intervalo desejado, estando abaixo ou acima do esperado, o MIVES avalia se existe a possibilidade de realizar a adequação métrica no segmento escandido (linhas 25 e 37, Algoritmo 1). Essa adequação métrica irá ajustar a escansão realizada sobre o segmento frásico, de modo que este fique dentro do metro esperado pelo o usuário. Os ajustes podem ser realizados da seguinte forma:

1. Caso a métrica alcançada pela escansão esteja acima do esperado: nestes casos o sistema avalia se reconsiderar as situações de diéreses aplicadas durante a escansão permitirá ter um verso dentro da métrica desejada. Tomemos como exemplo o segmento: “Varada a estreita faixa de cerrados,” este será escandido inicialmente pelo MIVES como “Va/ra/da a es/tre/i/ta/ fa/i/xa/ de/ce/rra/dos,” um dodecassílabo, por considerar dentre outros fenômenos, diérese nas palavras “estreita” e “faixa” que possuem forma de separação gramatical, respectivamente, como “es-trei-ta” e “fai-xa”. Se nessa situação, o metro esperado for um decassílabo, o sistema irá reconsiderar as diéreses aplicadas sobre estas palavras tornando o segmento um decassílabo “Va/ra/da a es/trei/ta/ fai/xa/ de/ce/rra/dos,”.
2. Caso a métrica alcançada pela escansão esteja abaixo do esperado: nestes casos, o sistema avalia se reconsiderar os fenômenos aplicados durante a escansão que causam uma diminuição do metro desejado (sinéreses, crases, elisões, sinalefas) irá fazer com que o segmento tenha o metro esperado pelo usuário. Por exemplo, esperando por um decassílabo, o segmento: “Hipóteses sobre a sua gênese. ”, será inicialmente escandido pelo MIVES como um octossílabo, “Hi/pó/te/ses/ so/bre a/ sua/ gê/ne/se.” Durante a escansão deste segmento, dois fenômenos que causam a diminuição do número de sílabas foram considerados pelo MIVES. A elisão entre a palavra “sobre” e o artigo “a” e a sinérese sobre a palavra “sua”. Neste caso o MIVES irá reconsiderar os casos de elisão e sinérese, ajustando o verso para que se torne um decassílabo, “Hi/pó/te/ses/ so/bre/a/ su/a/ gê/ne/se.”.

Depois que encontra uma forma de escansão, dentro dos metros informados pelo o usuário, o MIVES ainda tenta encontrar outras formações, ou possibilidades de escansão, para o mesmo segmento frásico utilizando as mesmas técnicas descritas até este ponto.

3.1.4. Unidade de Estrutura Métrica

A busca por estruturas de versificação pode se dar de três formas diferentes:

- Em frases completas: segmentos frásicos que em sua formação completa esteja dentro da métrica indicada pelo usuário;
- No início de Frases: Escansão feita no início dos segmentos frásicos;
- No final da Frase: Escansão realizada no conjunto de palavras que constituem o final de um segmento frásico.

Situações que exemplificam como essas estruturas de versificação podem ocorrer, podem ser vistos no trabalho realizado por [Almeida 1946] e por [Campos 1996], conforme apresentado na seção 2.2. A realização da busca em três diferentes regiões do texto, embora as características da escansão sejam as mesmas, exigem procedimentos internos distintos. A seguir uma descrição de como essas atividades são executadas internamente pelo sistema.

Frases Completas: Para realizar a escansão e identificação em segmentos frásicos completos, o MIVES processa a lista de segmentos frásicos extraídas do texto. A cada palavra processada, o sistema verifica se o segmento escandido até aquele ponto está entre os metros (contagem total de sílabas poéticas) máximo e mínimo configurados pelo usuário. Quando o padrão normatizado é encontrado, o sistema aborta a escansão daquele segmento, e verifica se dentro daquela estrutura de palavras escandidas constam todas as palavras do segmento frásico processado. Em caso positivo, a estrutura escandida é incluída na lista de resultados.

Como exemplo, vamos assumir que ao utilizar o sistema o usuário tenha parametrizado o *software* para realizar a busca por estruturas de versificação de frases completas, com metros entre 10 a 12 sílabas. Ao escandir “É o homem permanentemente fatigado.”, o sistema verifica se a quantidade de sílabas poéticas encontradas está dentro da faixa de metros solicitada, em seguida, verifica se a forma escandida possui o mesmo número de palavras contidas no segmento frásico. Caso seja verdadeiro, o segmento é incluído nos resultados. Caso seja falso, o sistema tentará outras variações métricas, que possam resultar em segmentos com estrutura métrica dentro dos metros informados pelo usuário.

É importante ressaltar que processos de escansão sobre o mesmo verso ou segmento frásico podem converter-se em resultados diferentes. Um mesmo segmento frásico pode ser escandido de diferentes formas, por esse motivo, o MIVES explora diferentes possibilidades métricas de escansão em um mesmo segmento frásico. Como exemplo, assumindo que o usuário deseja encontrar metros entre 10 e 12 sílabas poéticas, o MIVES apresenta três possibilidades de escansão para o segmento frásico “Rarefazem-se as matas, ou empobrecem” escandido pelo MIVES como decassílabo (Ra/re/fa/zem/se as/ ma/tas,/ ou em/po/bre/cem.), hendecassílabo (Ra/re/fa/zem/se/ as/ ma/tas,/ ou em/po/bre/cem.) ou dodecassílabo (Ra/re/fa/zem/se/ as/ ma/tas,/ ou em/po/bre/cem.).

Início de Frases: Neste tipo de busca, o MIVES tem um método de escansão semelhante a busca por estruturas de versificação de frases completas, porém além de utilizar como critério de parada o número de sílabas poéticas encontradas, o sistema utiliza como regra de validação do segmento frásico escandido, a sua delimitação por símbolos de pontuação (e.g. vírgula, exclamação, ponto e vírgula, dois pontos). Isso se fez importante pelo fato do MIVES lidar com aspectos métricos, outros instrumentos usados na escansão, como análise semântica ou fonética, são necessários para que não haja uma explosão do número de estruturas retornadas pelo MIVES. Outra questão importante é que, ao passo que a estrutura é escandida e atenda aos critérios descritos, esta é armazenada como resultado e a escansão no segmento frásico continua na tentativa de encontrar outras formações métricas. Isto é, quando o sistema encontra, no verso escandido, uma quantidade de sílabas dentro da faixa estabelecida pelo usuário, ele armazena a estrutura de versificação, com as palavras até aquele momento

escandidas, dentre os resultados encontrados e continua o processo de escansão, em busca de outras formações, desde que atendam a métrica. O processo é abortado quando a escansão alcança um número de sílabas superior a métrica indicada pelo usuário.

Para exemplificar, mais uma vez assumindo que os metros desejados pelo usuário estejam entre 10 e 12 sílabas, usaremos a estrutura de versificação "Varada a estreita faixa de cerrados," extraído de Os Sertões, pelo MIVES. Essa estrutura de versificação foi escandida a partir do segmento frásico:

“Varada a estreita faixa de cerrados, que perlongam aquele último rio, está-se em pleno agreste, no dizer expressivo dos matutos: arbúsculos quase sem pega sobre a terra escassa, enredados de esgalhos de onde irrompem, solitários, cereus rígidos e salientes, dando ao conjunto a aparência de uma margem de desertos”

Aqui, o sistema apresentou três possibilidades de escansão diferentes: a primeira como decassílabo “Va/rã/dá a es/trie/esta/ fia/xá/ de/ cem/rara/dos,” a segunda como um hendecassílabo “Va/rã/dá a es/trie/esta/ fá/i/xá/ de/ cem/rra/dos,” e a última como um dodecassílabo, “Va/ra/da a es/tre/i/ta/ fa/i/xa/ de/ ce/rra/dos,”. Neste caso, o sistema abortou o processo de escansão depois de duas conclusões: (a) continuar a escansão, processando as palavras “que” e “perlongam” resultaria em um metro acima do valor máximo, 12 sílabas informado pelo usuário e (b) depois da palavra “perlongam” não existe símbolo de pontuação que sirva como delimitador de sentença.

Final de Frase: Da mesma forma que descrita para buscas de estruturas de versificação de frases completas e para aquelas de início, o principal critério de parada é o número de sílabas poéticas e a regra de validação é estar entre os mesmos símbolos de pontuação utilizados como critério para buscar por estruturas de início de frase. Porém, em busca de reduzir o custo computacional e conseqüentemente o tempo de processamento, a escansão é iniciada a partir da última palavra em direção ao início do segmento frásico. Aqui, só interessa a estrutura de versificação formada pelas últimas palavras do segmento frásico. Portanto, a estratégia de realizar o processo da última em direção a primeira palavra, permite que, em segmentos frásicos extensos, o sistema não tenha custo computacional processando palavras que, já se sabe, não farão parte dos resultados. Vejamos o exemplo da estrutura de versificação, “o aspecto atormentado das paisagens. ”, essa estrutura de versificação foi extraída de Os Sertões, pelo MIVES a partir do segmento frásico:

“As condições estruturais da terra lá se vincularam à violência máxima dos agentes exteriores para o desenho de relevos estupendos. O régimen torrencial dos climas excessivos, sobrevindo, de súbito, depois das insolações demoradas, e embatendo naqueles pendores, expôs há muito, arrebatando-lhes para longe todos os elementos degradados, as séries mais antigas daqueles últimos rebentos das montanhas: todas as variedades cristalinas, e os quartzitos ásperos, e as filades e calcários, revezando-se ou entrelaçando-se, repontando duramente a cada passo, mal cobertos por uma flora tolhiça — dispondo-se em cenários em que ressalta predominante, o aspecto atormentado das paisagens”.

Realizar o processamento de escansão neste segmento frásico da primeira para a última palavra apenas exigiria custo computacional desnecessário e conseqüentemente mais tempo de processamento. Aqui também tomamos como base que o usuário tenha utilizado como metros, extremos entre 10 e 12 sílabas poéticas. Para este caso, o MIVES

apresentou a estrutura de versificação em um decassílabo escandido como “o as/pec/to a/tor/men/ta/do/ das/ pa/i/sa/gens.”. Observando o segmento frásico de origem da estrutura de versificação, pode-se observar que seria desnecessário incluir na escansão a palavra “predominante”, o que resultaria em um metro fora dos limites pelos usuários.

3.2 Visualização e Análise dos Resultados

A visualização dos resultados tem como objetivo, prover mecanismos capazes de apresentar visualmente os dados aos usuários, permitindo que estes possam organizar, classificar, analisar dados e priorizar informações de forma intuitiva.

Por ser um campo ainda inexplorado pela comunidade científica, a visualização dos resultados e análise do tipo de informação entregue pelo MIVES, implementamos mecanismos, identificados na pesquisa como necessários, de modo que, o usuário tenha a possibilidade de obter conclusões sobre os resultados de forma consistente, através do uso de ferramentas originalmente pensadas para facilitar a interpretação deste tipo de informação.

Priorizamos uma visão contextualizada da informação. O usuário tem acesso ao documento originalmente submetido ao MIVES com as estruturas de versificação subliminares sobreposta ao texto. Mantendo a originalidade do documento e permitindo uma visualização contextualizada do verso apresentado. É extremamente importante deixar claro que, analisar contextualmente uma estrutura de versificação é fundamental para a plena visualização do fenômeno ocorrido e as derivações impostas pelo ambiente ao qual este se apresenta.

Na maioria dos casos, as estruturas de versificação identificadas pelo MIVES, apresentam-se distribuídas ao longo do documento. Algumas relativamente distantes uma das outras. Em obras literárias muito extensas e com longas distâncias entre as estruturas, a busca pode se tornar cansativa, inviabilizar a análise e, conseqüentemente a priorização das informações relevantes. Para evitar esse tipo de cenário, implementamos mecanismos de navegabilidade que permitem o trânsito sequencial sobre as estruturas apresentadas no próprio texto ou, caso seja preferido pelo usuário, desprendidas deste. Em ambos os casos, durante a navegação, usuário tem a possibilidade de visualizar informações sobre cada estrutura apresentada, como: i) o verso em sua forma escandida; ii) o posicionamento das tônicas; iii) a classificação ou as diferentes classificações métricas sobre um mesmo verso; iv) a quantidade, quais foram e em que palavra (s) do verso foram considerados os fenômenos Intravocabulares e; v) quantos fenômenos Intervocabulares foram considerados naquele verso durante a escansão.

Além da análise das estruturas de versificação, o MIVES provê gráficos e dados estatísticos. Estes dados, permitem ao usuário ter informações a respeito das frequências das estruturas de versificação dentro um intervalo de segmentos frásicos. Esse intervalo pode ser facilmente ajustado pelo usuário. As distâncias existentes entre os segmentos frásicos também podem ser visualmente analisadas em um gráfico de distância disponibilizado pela ferramenta. Apesar disto, desenvolver um mecanismo de análise capaz de contemplar todas as possibilidades e perspectivas de visualização sobre um determinado volume de dados, não é uma tarefa trivial. Diante da impossibilidade de

resolver essa questão, não limitando as análises dos usuários aos recursos disponíveis no sistema. O MIVES permite que os resultados obtidos, como por exemplo as estruturas de versificação, os valores de distâncias, os segmentos frásicos de origem com suas respectivas informações de metros encontrados, dentre outros, possam ser exportados. Dessa forma, o usuário tem a possibilidade de utilizar outras ferramentas para visualizar e manipular a informação da forma que desejar.

3.3 Sistema MIVES

A arquitetura do MIVES possui módulos bem definidos, com tarefas específicas que interagem, conforme apresentado na Figura 1. Vejamos, a seguir, uma descrição destes módulos.

Módulo de Extração de Segmentos: divide o texto em segmentos frásicos. O resultado deste processamento é entregue ao módulo de escansão;

Módulo de Escansão e Identificação das Estruturas de Versificação: realiza o processamento das estruturas frásicas fornecidas pelo Módulo de Extração de Segmentos. O Módulo de Escansão representa o núcleo da aplicação. Todos os mapeamentos entre segmentos frásicos da obra e as estruturas de versificação são realizadas por este módulo. Ele faz uso do módulo de separação silábica e produz os resultados que servem como entrada para o módulo de Análise e Visualização de Resultados.

Módulo de Separação Silábica: Este módulo é constituído por dois elementos básicos, um dicionário, baseado em uma estrutura em *hash*, e um algoritmo que realiza a separação silábica inicial e identificação de tônicas. A separação é inicial porque a forma assumida pela palavra separada poderá sofrer alteração pelo módulo de escansão.

A construção deste módulo tem o objetivo de economizar recursos computacionais e consequentemente tempo de processamento. Conforme exemplo apresentado na Tabela 4, o dicionário do MIVES possui entradas que representam, em sua primeira coluna a palavra em sua grafia original, seguida de sua forma em separação gramatical conforme retorno do *LapSeparator*. Quando o Módulo de Escansão, solicita que uma palavra seja separada, é acionado o Dicionário e, duas situações podem ocorrer: (a) caso a palavra esteja dentre suas entradas, é retornado diretamente pelo dicionário a separação gramatical daquela palavra ao Módulo de Escansão; (b) caso não conste entre as entradas do dicionário, o *LapSeparator* é acionado, o retorno deste é inserido no dicionário e, em seguida, o dicionário retorna ao Módulo de Escansão.

Conforme descrito no parágrafo anterior, quando uma palavra não consta no dicionário é acionado o *LapSeparator*. Este é o algoritmo de separação silábica proposto por Neto, [2015], conforme descrito na seção 3.1.4, foi escrito em uma linguagem de programação diferente daquela utilizada inicialmente em nossa pesquisa. Porque não há interoperabilidade direta entre o Java, utilizado em nosso trabalho, e o *CSharp*, utilizado no algoritmo de separação, é utilizado o *JNI4NET* para permitir que o código Java tenha acesso ao código nativo escrito em *CSharp*. O *JNI4NET* funciona como uma ponte, e permite a comunicação entre os dois códigos. Mais detalhes a respeito do *JNI4NET* podem ser encontrados na página do projeto: <http://jni4net.com>.

Tabela 4: Exemplo de Conteúdo de Dicionário

Chave	Valor
cunha	c#u-nha
nota	n#o-ta
preliminar	pre-li-mi-n#ar
escrito	es-cr#i-to
nos	nos
raros	r#a-ros
intervalos	in-ter-v#a-los
de	de
folga	f#ol-ga
uma	#u-ma
demos-lhe	d#e-mos§-lhe
isto	#is-to
outra	#ou-tra
feição	fei-ç#ão

O Dicionário baseado em *hash*, permite um tempo de recuperação na ordem de 1, $O(1)$. Houve perceptiva diminuição no tempo de processamento de textos em que as palavras constavam no dicionário. Por meio dessa estratégia, a medida que novas obras são processadas, o sistema tende a ficar mais rápido, já que o dicionário se torna mais rico, diminuindo gradativamente a necessidade de processar palavras pelo acionamento do *LapSeparator*. Embora a integração entre o nosso código, escrito em Java e o *LapSeparator*, escrito em *CSharp*, tenha resolvido a questão da separação silábica, produziu uma dependência direta com a plataforma operacional a qual depende o código nativo do *LapSeparator*.

Módulo de Visualização dos Resultados: Para a identificação de estruturas subliminares de versificação, por ser uma área de pesquisa inexplorada, não encontramos exemplos de sistemas que tratam da visualização e análise de dados desta natureza. Neste módulo, são fornecidos os mecanismos, até o momento, identificados em nossos experimentos como necessários. As funcionalidades incorporadas a este módulo foram apresentadas na seção 3.2. Os resultados da utilização prática serão apresentados nesta seção, Interface de Visualização e Análise de Resultados.

Método de Utilização do MIVES

Com o objetivo de criar uma ferramenta com rotinas que pudessem ser facilmente aprendidas pelo usuário, todo o processo que antecede a análise de resultados funciona como um passo-a-passo conforme apresentado há Figura 2.

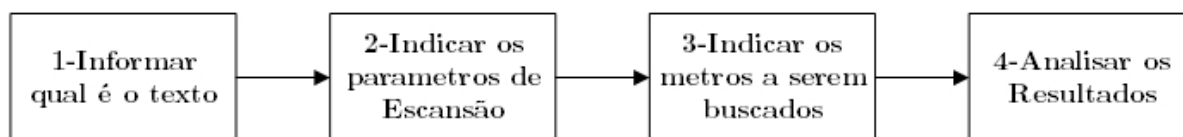


Figura 2: Passos para utilização do MIVES

O primeiro passo é indicar qual texto deve ser processado (Figura 3) seguido pelos parâmetros que devem ser considerados durante a escansão (Figura 4) até os metros e quais locais devem ser buscados (Figura 5), são configurados através de interfaces simples e intuitivas.

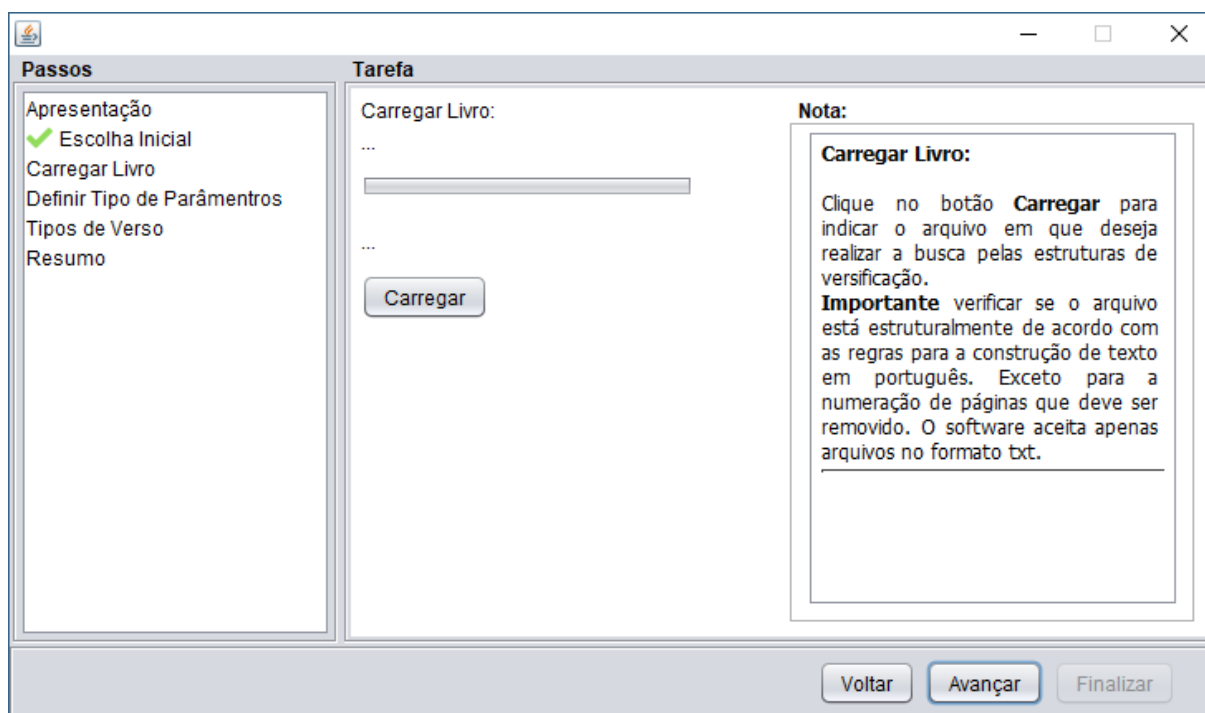


Figura 3: Interface para indicar a arquivo a ser processado pelo MIVES

A Figura 3, apresenta a interface fornecida pelo MIVES para que o usuário indique qual deverá ser o texto a ser processado.

A Figura 4, apresenta a interface que possibilita ao usuário configurar seus próprios parâmetros de escansão para os fenômenos fonológicos inerentes a escansão (diéreses sinéreses, sinalefas, elisão e crase). A configuração destes parâmetros é feita em painéis tabulados que agrupam configurações de fenômenos fonológicos específicos, ou seja, para cada tipo de fenômeno fonológico é utilizado um painel específico, no qual podem ser inseridos os pares de vogais, conforme descrito na seção 3.1.3, e como deve ser o comportamento do MIVES mediante a ocorrência deste par de vogais durante o

processo de escansão.

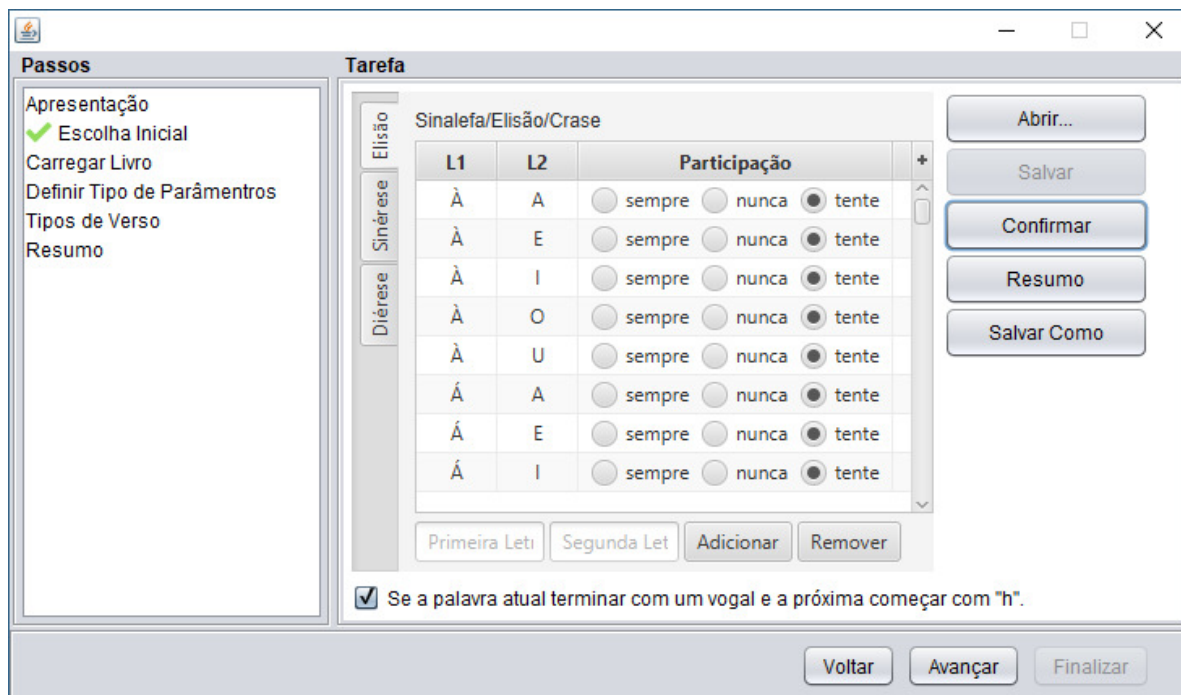


Figura 4: Interface de configuração dos parâmetros de escansão do MIVES

Antes que o processo de mineração pelas estruturas de versificação seja iniciado, o usuário deve informar através da interface apresentada pela Figura 5, quais são os metros mínimo e máximo que devem ser considerados, bem como o tipo de busca (sentenças completas, início ou final) que deve ser feito nas sentenças do texto.

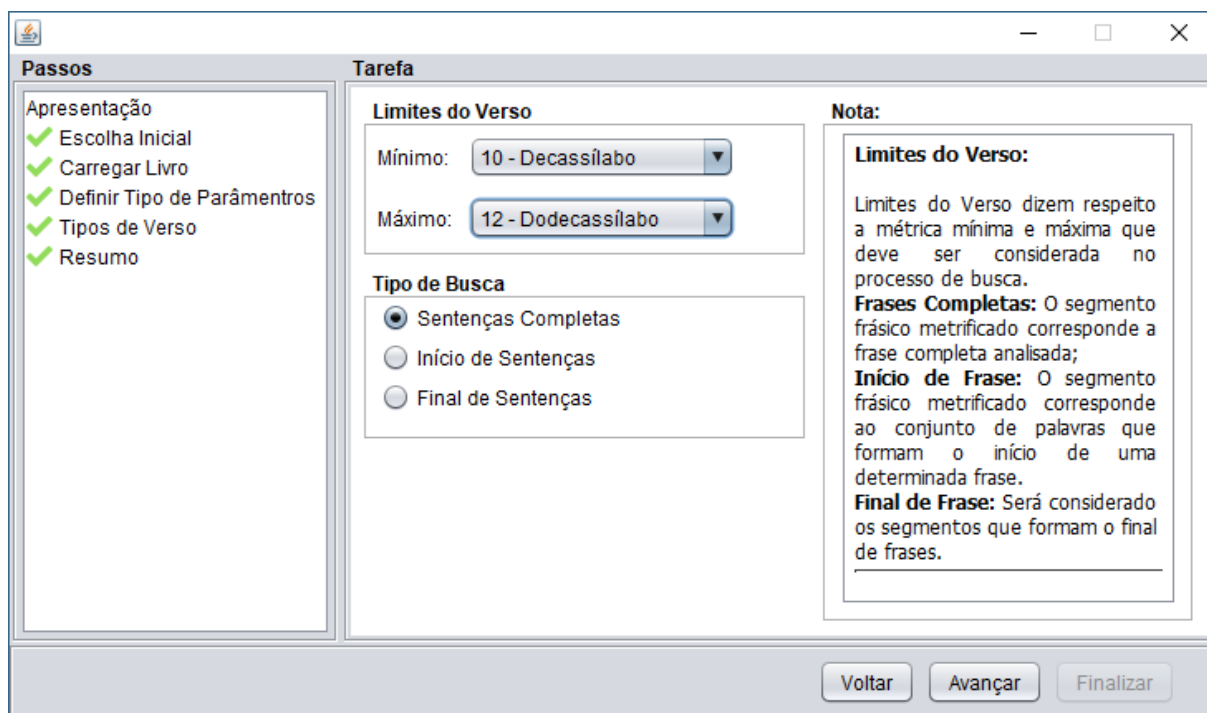


Figura 5: Interface de Configuração de Metros e Tipo de Busca

A etapa de visualização (ver Figura 6) dos resultados é disponibilizada ao usuário logo depois que o MIVES conclui o processo de busca.

Interface de Visualização e Análise dos Resultados

Concebida a partir da necessidade de analisar os textos processados pelo módulo de escansão, a interface de visualização permite que os resultados sejam analisados de forma contextualizada. A contextualização anunciada aqui diz respeito a capacidade do sistema de realçar as estruturas identificadas no texto, permitindo que o usuário realize a análise do segmento escandido no local onde este foi escrito pelo autor da obra, conforme pode ser observado na Figura 6 (em 1 – Área do Texto).

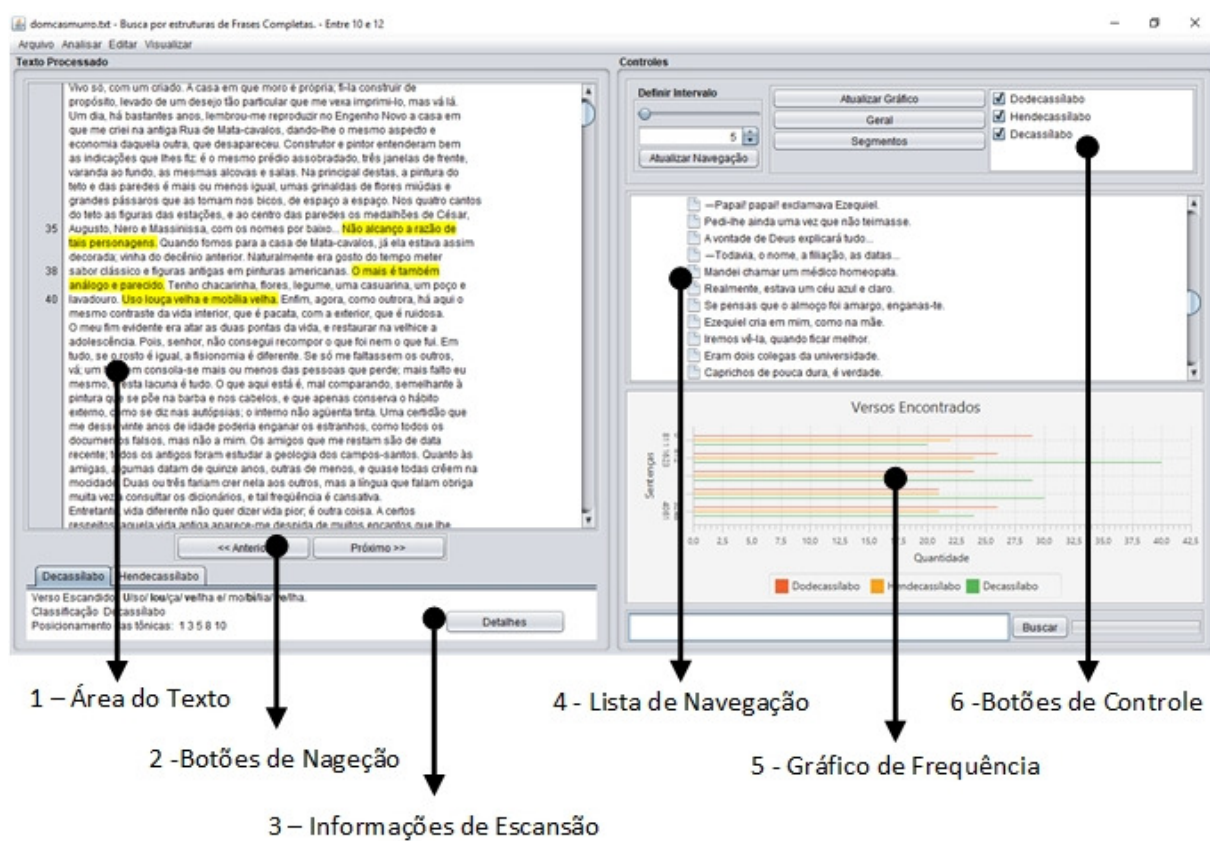


Figura 6: Interface de Visualização e Análise dos Resultados

A interface do MIVES possui controles que funcionam de forma integrada permitindo que o usuário tenha a possibilidade de analisar os resultados navegando por estes de diferentes formas. A área de texto, os botões de navegação, o painel com as informações de escansão, árvore de navegação, os botões de controle e o gráfico de frequência absoluta são recursos de análise que funcionam de forma integrada. Enquanto a Área de Texto permite que o usuário tenha acesso ao documento submetido ao MIVES, com as sentenças e a estrutura métrica destacadas (em amarelo na Figura 6). Os botões de navegação permitem que o usuário possa navegar sequencialmente por entre as sentenças identificadas, ao passo que essa navegação é executada, o painel com detalhes acompanha, apresentando informações da escansão, ou as diferentes formas de escansão,

sobre o segmento selecionado. Além do painel de detalhes (Figura 6 em 3 – Informações de Escansão), acompanha o processo da Lista de Navegação, destacando dentre as estruturas apresentadas por ela, aquela que atualmente está sendo analisada pelo usuário. Essa lista de navegação, permite que o usuário escolha aleatoriamente, a estrutura de versificação que deverá ser analisada, sendo essa seleção sincronizada com a área de texto, os botões de navegação, e o painel de detalhes.

Os botões de controle definem como as informações na lista de navegação e no gráfico serão organizadas. Estes são utilizados para dividir a análise do texto por intervalo de sentenças. Agrupando sentenças com estrutura métrica pertencentes a aquele intervalo em grupos de análise na lista de navegação e atualizando o gráfico com as frequências absolutas por intervalo.

Além das informações apresentadas pela interface principal de análise (Figura 6) o usuário ainda conta com uma tela que permite navegar apenas dentre os resultados, tendo informações mais amplas sobre cada segmento escandido, como os possíveis números de sinéreses, diéreses e elisões consideradas sobre ao segmento em destaque, por metro escandido, conforme pode ser observado na Figura 7.

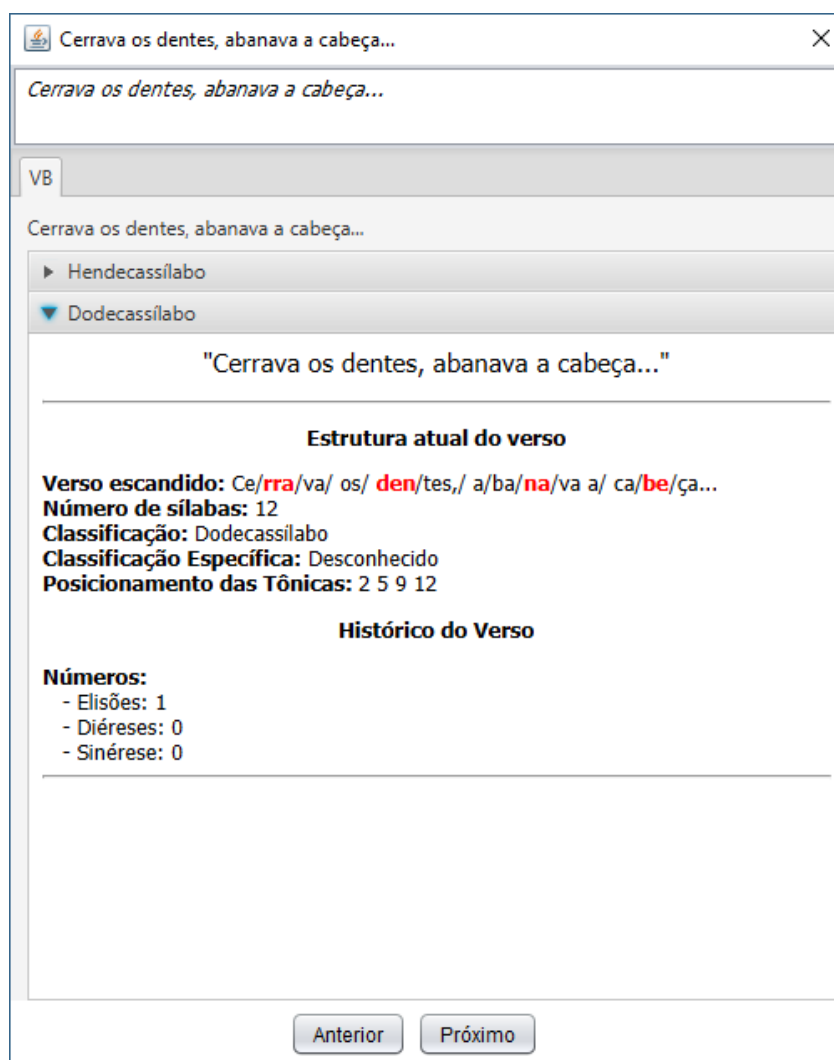


Figura 7: Interface de Detalhes do Segmento Escandido.

Diante da necessidade de uma análise mais detalhada do gráfico apresentado pela interface de análise (Figura 6, em 5 – Gráfico de Frequência), o usuário pode acessá-lo através do menu, Analisar, apresentado na Figura 6 e ter uma visão mais ampla, conforme apresentado na Figura 8. Outro gráfico também disponível no MIVES é o gráfico que apresenta as distâncias entre as estruturas de versificação identificadas pelo sistema.

Além disso, através desta interface (Figura 6) é possível exportar os resultados obtidos para arquivos devidamente organizados, permitindo a importação destes por meio de outra ferramenta de análise, preferida pelo usuário.

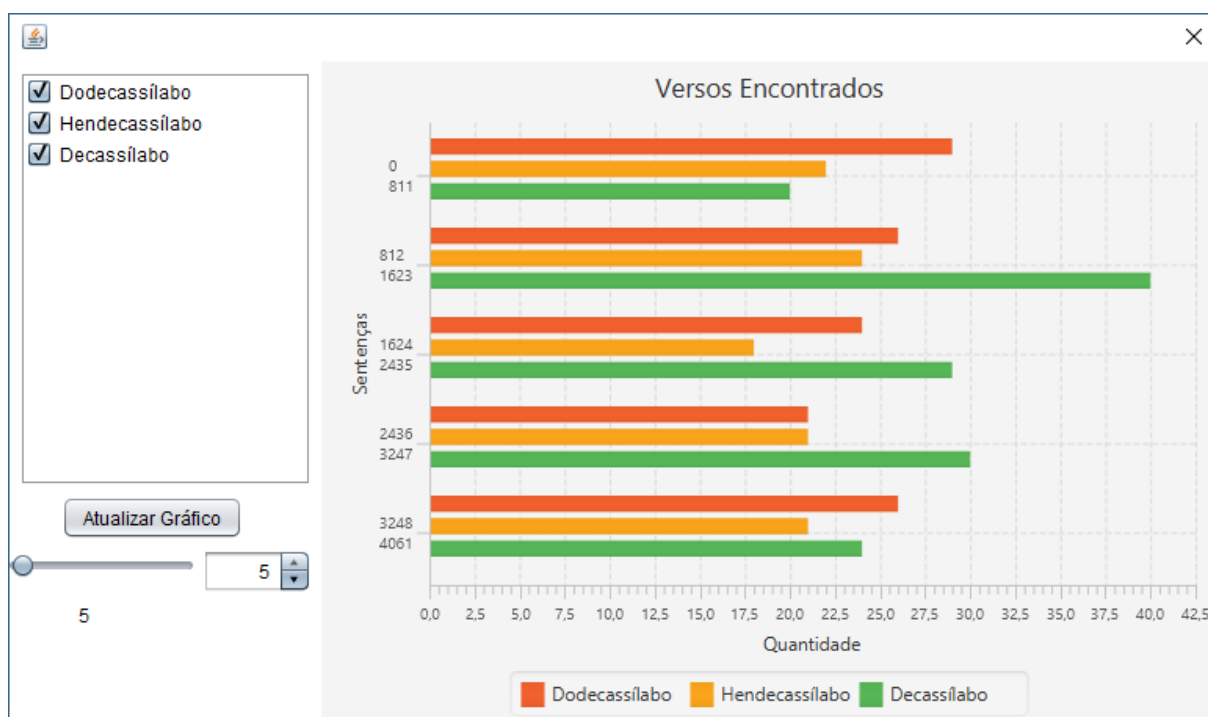


Figura 8: Gráficos de Frequência Absoluta do MIVES

Existem outras capacidades de análises disponíveis, destacamos aqui as que consideramos mais importantes segundo aquilo que foi exigido por nossas experiências de análise em laboratório. De acordo com nossos experimentos e análises de resultados, a proposta desta interface se apresenta como um recurso que se mostrou útil e merece destaque no MIVES.

Capítulo 4

Resultados

Retomando, o objetivo principal desta pesquisa é propor uma metodologia computacional para identificação automática de estruturas de versificação na prosa literária brasileira. Para validar nossa proposta construímos o MIVES, uma ferramenta computacional que implementa a metodologia proposta e fornece uma interface de análise para os resultados. Esta seção está dedicada a apresentar os resultados oriundos da aplicação do MIVES sobre textos da prosa literária brasileira.

Para todos os experimentos apresentados neste capítulo, utilizamos o conjunto de parâmetros de escansão apresentados no Anexo 1. Entende-se por parâmetro de escansão um par de vogais associado a um tipo de fenômeno fonológico. Este conjunto de parâmetros foi gerado a partir da combinação entre 29 vogais extraídos da tabela ASCII (*American Standard Code for Information Interchange*) produzindo um total de 841 pares de vogais, que foram vinculados a cada um dos tipos de fenômenos fonológicos tratados pelo MIVES. Além disso, o MIVES foi configurado para tentar aplicar o fenômeno fonológico, conforme descrito na seção 3.1 Etapas do MIVES. Optamos por utilizar todas as combinações de encontros entre vogais para que nossas preferências não interferissem nos resultados obtidos ao fim de cada experimento.

Dividimos esta seção em duas partes. Na primeira, validamos o processo de escansão utilizando uma obra com estruturas métricas já conhecidas. Já na segunda etapa, apresentamos os resultados relacionados com o núcleo da aplicação, ou seja, o processo de escansão e identificação das estruturas de versificação.

4.1 Validação do Processo de Escansão

Conforme já descrito em nossa metodologia, a busca por estruturas de versificação envolve um variado conjunto de processos. Dentre os processos apresentados, a correta escansão dos segmentos frásicos é chave para identificar as estruturas subliminares de versificação. A motivação em validar o processo de escansão realizado pelo MIVES se fez importante para respaldar os resultados obtidos na escansão dos segmentos de texto, oriundos da prosa apresentados pelo MIVES, como possíveis estruturas de versificação. A falta de trabalhos relacionados diretamente a essa pesquisa nos obrigou a buscar por obras poéticas, em que já se conheciam, desde já, a estrutura métrica dos versos.

Para esse processo, escolhemos a obra de Luíz Vaz de Camões, *Os Lusíadas*. Publicada

em 1572, é conhecida como uma obra prima da poesia portuguesa, dentre outros elementos, por ser constituída de 8.816 versos, todos decassilábicos. Por ser uma obra com estrutura de versos e estrofes, foi necessário realizar uma adequação no formato do texto para que o MIVES considerasse cada verso do poema como um segmento frásico a ser escandido. Esta ação foi baseada na substituição de todos os símbolos de pontuação no final de cada verso, por um ponto de final de sentença, sendo inserido também, nos versos que não tinham símbolos de pontuação. Esta ação se fez necessária devido a etapa de extração de segmentos frásicos, executada pelo MIVES, no texto em processamento.

Buscamos inicialmente apenas por decassílabos, ao final do processo, 8256 sentenças foram escandidas como versos decassilábicos, correspondendo a uma taxa de identificação de cerca de 94%. Para testar outras possibilidades de escansão sobre os mesmos versos Camões, repetimos o experimento desta vez utilizando como restrição métrica de 10 a 12 sílabas poéticas. Entre decassílabos a dodecassílabos, o MIVES retornou 8779 versos escandidos. Além de escandir 82565 versos como decassílabos, o MIVES conseguiu variações de escansão a ponto de reportar 6730 hendecassílabos e 3721 como dodecassílabos.

Voltando apenas a tratar dos decassílabos, cerca de 6% ou 560 versos não foram escandidos conforme esperava-se. Não houve tempo suficiente para analisar todos os aspectos que levaram o MIVES a não escandir os 100% dos versos como decassílabos. Dentre os problemas identificados estão os casos de Eclipse, erros do separador silábico e situações de uso de alguns diacríticos, em palavras presentes na obra utilizada no experimento que já não são utilizadas na escrita contemporânea. A Eclipse é um fenômeno fonológico não considerado pelo sistema. É caracterizado pela elisão da consoante “m”, final de uma palavra, com a inicial vogal da palavra seguinte. “A esta fusão vocálica, facilitada pela perda da ressonância nasal da primeira vogal, dar-se o nome de Eclipse” [Cunha e Cintra, 2008, p. 688]. É o que ocorre com o verso “Dei/**xou**,/ com/ a/ **gran**/de e/ **prós**/pe/ra/ vi/**tó**/ria.”, escandido pelo MIVES com um hendecassílabo, poderia se tornar um decassílabo caso fosse considerada elisão entre a palavra “com” e a vogal “a”. Os casos de Eclipse não foram considerados até esta etapa dos nossos estudos por serem tão comuns. Este fato consta no estudo apresentado por [Mittmann, 2016, p 139], quando demonstra em seu trabalho a quantidade de aplicações das regras (fenômenos fonológicos) por 10 mil sílabas, agrupadas por categoria.

Além dos casos de Eclipse, identificamos alguns equívocos de identificação de tônicas realizado pelo *LapSeparator*, Neto[2015]. Um destes casos de erro na identificação da tônica ocorreu com a palavra “triunfando” do verso “De nações diferentes triunfando.” Escandido pelo MIVES como um verso de 8 sílabas poéticas (De/ na/**ç**ões/ di/fe/**ren**/tes/ **tri**/un/fan/do). Nesta escansão, pôde ser observado que a sílaba tônica da palavra “triunfando” foi indicada incorretamente na primeira sílaba da palavra, fazendo o MIVES desprezar duas sílabas poéticas e retomando o verso como um octossílabo. Outro ponto relacionado a equívocos cometidos pelo *LapSeparator*, está relacionado a identificação de tônicas em palavras em que a grafia atual já despreza uso de alguns diacríticos. Isso foi observado na palavra “sàbiamente”, no verso “Guiados pelas ondas sàbiamente.”, o diacrítico utilizado por Camões na vogal “a” da primeira sílaba da palavra, fez com que o *LapSeparator* indicasse como tônica a primeira sílaba

da palavra. Neste último caso, o MIVES também desprezou dentre os resultados, este verso, por ter sido escandido apenas como 8 sílabas poéticas. Felizmente a capacidade fornecida pelo MIVES de permitir que o usuário realize intervenções no dicionário, nos permitiu corrigir, nestas palavras, o equívoco cometido pelo *LapSeparator* e, estes versos foram incluídos dentre os resultados.

Apesar das limitações identificadas pelo processo de escansão, os resultados obtidos com a escansão de *Os Lusíadas*, chegaram a cerca de 94% rendimento.

4.2 Processo de Escansão e Identificação das Estruturas de Versificação

Conforme descrito na seção **Metodologia**, essa pesquisa foi, inicialmente, inspirada nos trabalhos realizados por Almeida [1946] e por Campos[1996]. Estes poetas indicaram a ocorrência de estruturas de versificação em diversas regiões do texto na obra de Euclides da Cunha, “*Os Sertões*”. Dentre os inúmeros tipos de formações apresentadas por Almeida [1946] e Campos [1996], se destacam as estruturas de (a) frase completa (Campos [1996] as definiu como “parágrafos destacados e autônomos”), (b) de início de frase e (c) aqueles que constituem o final de alguns segmentos de texto, classificados por Campos [1996] como sintagmas conclusivos de orações. Para expandir nossas análises, além compararmos os resultados obtidos pelo MIVES em os “*Os Sertões*” com aqueles apresentados nos trabalhos de Almeida[1946] e Campos[1996], executamos o mesmo procedimento em outras obras de Euclides da Cunha, e publicações de três autores do mesmo período. A motivação para fazermos os experimentos com outras obras de Euclides foi o de verificar se o comportamento versificatório presente em *Os Sertões* se repetia em outras obras do mesmo autor. Já a análise realizada em outras obras de diferentes autores teve como ambição analisar se o comportamento identificado na escrita de Euclides se repetiria em outras obras de diferentes autores do mesmo período literário do Brasil.

4.2.1 Análise de *Os Sertões*

[Campos 1996] encontrou mais de quinhentos decassílabos e mais de duas centenas de dodecassílabos na prosa de *Os Sertões*. Entretanto, nem ele, nem Almeida [1946], apresentaram, em seus artigos, uma lista completa das estruturas de versificação encontradas, nem revelaram as quantidades exatas de seus achados, apresentando apenas alguns exemplos. Baseado nestes exemplos, fizemos nossas primeiras experiências considerando como elemento qualitativo a equivalência de resultados entre as estruturas de versificação apresentadas pelo sistema com os achados de Campos. Optamos em trabalhar com os versos apresentados por Campos por que os mesmos incorporam aqueles apresentados por Almeida em seu artigo.

Sabendo da preferência de Euclides da Cunha por decassílabos e dodecassílabos, o experimento foi realizado buscando estruturas que foram divididas em três partes tais como:

Na primeira, busca-se por estruturas de versificação formadas por frases completas (parágrafos autônomos). Na segunda etapa, buscamos por estruturas de versificação de início e, por último, na terceira, pelos finais de sentenças. Ao todo, foram indicadas pelo software 4.287 estruturas de versificação, sendo: 652 de frases completas, 1753 de estruturas de início de frase e, 1862 dentre os segmentos de texto conclusivos de oração.

Etapa 1: Início de sentenças com Estrutura Métrica.

Na Tabela 5 são apresentados, na primeira coluna, os exemplos de decassílabos ou dodecassílabos listados por Campos [1996] e, nas duas colunas seguintes, respectivamente, as estruturas escandidas e as classificações métricas atribuídas pelo MIVES.

Tabela 5: Relação de Estruturas de Versificação Etapa 1 - Primeiro experimento

Transertões	MIVES	Nº Sílabas Poéticas
É o homem permanentemente fatigado.	É o ho /mem/ per/ma/nen/te/ men /te/ fa/ti/ ga /do.	11
	É /o ho /mem/ per/ma/nen/te/ men /te/ fa/ti/ ga /do.	12
É uma paragem impressionada.	É u /ma/ pa/ ra /gem/ im/pre/ssio/na/ do /ra.	10
	É/ u /ma/ pa/ ra /gem/ im/pre/ssio/na/ do /ra.	11
	É/ u /ma/ pa/ ra /gem/ im/pre/ssi/o/na/ do /ra.	12
É uma mutação de apoteose.	É/ u /ma/ mu/ta/ ção / de a/po/te/o/se.	10
Entra-se, de surpresa, no deserto.	En /tra-/se,/ de/ sur/ pre /sa,/ no/ de/ ser /to.	10
É a escarpa abrupta e viva dos planaltos.	É a es/ car /pa a/ brup /ta e/ vi /va/ dos/ pla/ nal /tos.	10
	É/ a es/ car /pa a/ brup /ta e/ vi /va/ dos/ pla/ nal /tos.	11
	É/ a/ es/ car /pa a/ brup /ta e/ vi /va/ dos/ pla/ nal /tos.	12
E assim passam numa agitação estéril.	E a/ ssim / pa /ssam/ nu /ma a/gi/ta/ ção es/ té /ril.	10

	E a/ ssim / pa /ssam/ nu /ma a/gi/ta/ ção / es/ té /ril.	11
É o prelúdio da sua desgraça.	É/ o/ pre/ lú /dio/ da/ su /a/ des/ gra /ça.	10
É mais um inimigo a suplantar.	É/ mais / um/ i/ni/ mi /go a/ su/plan/ tar .	10
	É/ mais / um/ i/ni/ mi /go/ a/ su/plan/ tar .	11
Esta ilusão é empolgante ao longe.	Es /ta i/lu/ são / é/ em/pol/ gan /te ao/ lon /ge.	10
	Es /ta/ i/lu/ são / é/ em/pol/ gan /te ao/ lon /ge.	11
	Es /ta/ i/lu/ são / é/ em/pol/ gan /te/ ao/ lon /ge.	12
O sertanejo é, antes de tudo, um forte.	O/ ser/ta/ ne /joé, an /tes/ de/ tu /do, um/ for /te.	10
	O/ ser/ta/ ne /jo é,/ an /tes/ de/ tu /do, um/ for /te.	11
	O/ ser/ta/ ne /jo/ é,/ an /tes/ de/ tu /do, um/ for /te.	12

Nesta etapa, o MIVES atingiu cerca de 97% de precisão ao incluir em seus resultados, como decassílabo ou dodecassílabo, 35 dos 36 versos, apresentados por [Campos 1996] formados por sentenças completas. A estrutura de versificação “Um primor de estatuária modelado em lama. ”, apresentada por campos como um dodecassílabo não foi incluída pelo MIVES em seus resultados por ter sido escandida com métrica superior a 12 sílabas poéticas, “Um/ pri/**mor**/ de es/ta/tu/**á**/ria/ mo/de/**la**/do em/ **la**/ma. ”. Neste caso, o sistema encontrou, desprezando a última sílaba átona, 13 sílabas poéticas. Dois fenômenos entre palavras foram considerados nesta sentença, entre as palavras “de” e “estatuária”, e entre as palavras “modelado” e “em”. Para tornar a sentença um dodecassílabo, em “estatuária”, precisaríamos considerar a possibilidade de aplicar sinérese entre a 3ª e 4ª sílaba desta palavra. Neste caso, o sistema não aplicou sinérese por que o MIVES sempre preserva a autonomia da sílaba tônica que sucede uma vogal. Esta diferença de métrica advém de variações possíveis na escansão para além das contempladas no MIVES.

O processo de escansão não é rígido e conseqüentemente não produz resultados inequívocos. Em uma mesma estrutura de versificação diferentes metros podem ser encontrados. Tendo como base esse princípio o MIVES busca realizar escansões diferentes sobre o mesmo segmento frásico. Isso pode ser observado na Tabela 5 quando, por exemplo na estrutura de versificação “Esta ilusão é empolgante ao longe. ”, o sistema

apresenta três variações de escansões possíveis. A quantidade de variações que o sistema tem como objetivo alcançar é controlado pelo o intervalo de metros, sobre as possíveis estruturas de versificação que o usuário espera encontrar. No caso da sentença “Esta ilusão é empolgante ao longe.” foram apresentadas três variações. Isso ocorre porque o metro que estávamos buscando pode variar entre 10 e 12 sílabas poéticas. Sendo assim, o sistema tentará encontrar 3 variações de escansão desde que a métrica esteja entre os limites de metros informados.

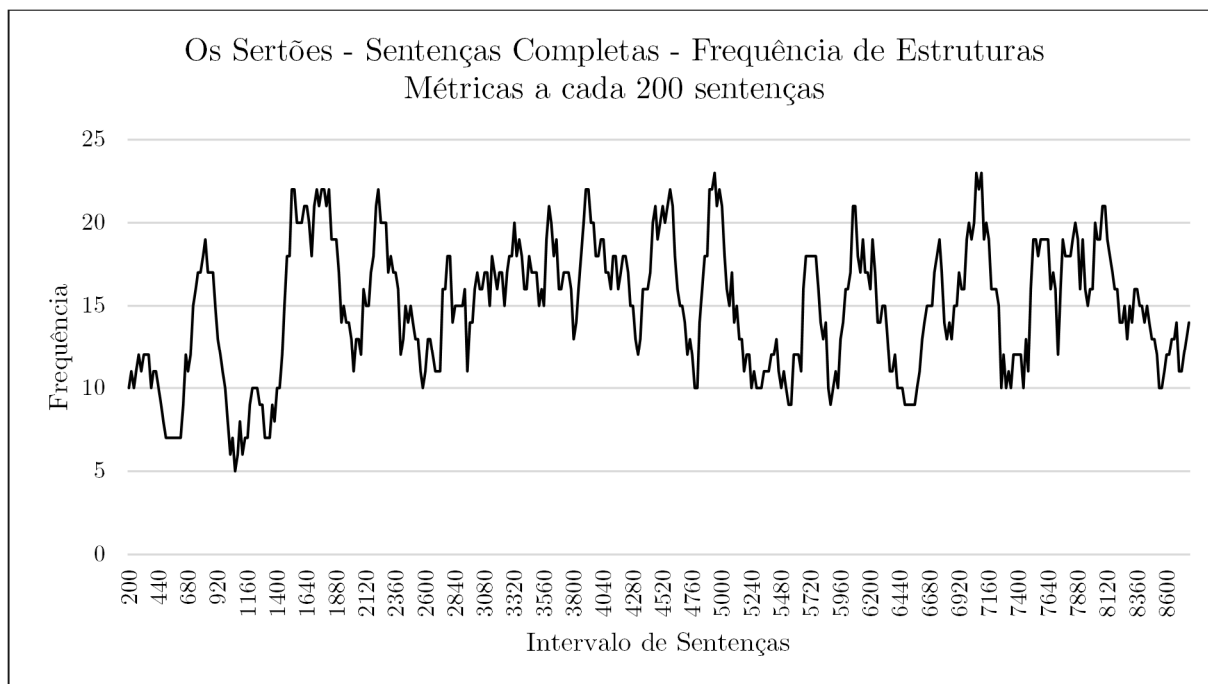


Figura 9: Os Sertões - Sentenças Completas - Distribuição da Frequência Absoluta.

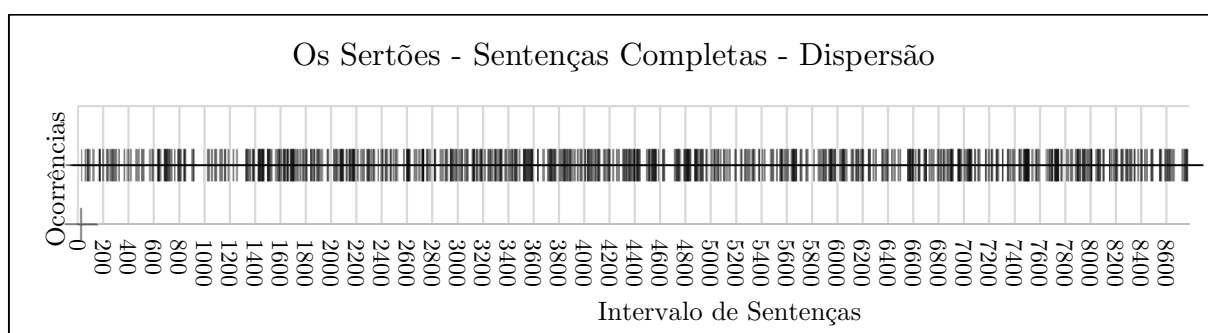


Figura 10: Os Sertões - Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto

O gráfico na Figura 9 apresenta a frequência absoluta das estruturas métricas agrupadas por intervalo de sentenças. Para todos os experimentos, analisamos estes dados dividindo todos os segmentos do texto em intervalos/janelas de 200 segmentos e um passo de 20 segmentos. Ou seja, é calculada a frequência absoluta entre os 200 primeiros segmentos (tamanho da janela) do texto e em seguida avançamos 20 segmentos (passo 20), construindo um novo intervalo. Dessa forma é possível visualizar como estão distribuídas as estruturas de versificação sem dividir em intervalos isolados os segmentos que fazem parte do livro. Na Figura 9, é possível ver no gráfico que existe

nesta obra, intervalos com maior densidade de estruturas métricas. Ou seja, há regiões no texto em que se tem uma concentração maior destas estruturas, é o que acontece, por exemplo, em algumas regiões do gráfico em que se percebe uma amplitude nos valores de frequência. Já o gráfico da Figura 10 apresenta como estas estruturas estão distribuídas ao longo do texto. Comparando as informações destes gráficos é possível identificar uma relação direta entre os dados que apresentam. Para simplificar, quando existe uma queda nos valores de frequência no gráfico da Figura 9 é possível perceber que na mesma região (intervalo) no gráfico da Figura 10 existe uma distância maior entre as sentenças com estruturas métricas identificadas.

Incluído 35 das 36 estruturas de versificação apresentadas por Campos, o sistema apresentou em seus resultados 652 sentenças com estrutura métrica. Dentre este decassílabos e dodecassílabos perfeitos, não apresentados por Augusto de Campos em seu artigo, como o decassílabo heroico “Assaltam-no; aferram-no; jugulam-no” escandido pelo MIVES como “A/ssal/tam/no;/ a/fe/rram/no;/ ju/gu/lam/no.” e pelo Alexandrino “Empanaram-se todas as vistas, de lágrimas...” escandido pelo MIVES como “Em/pa/na/ram/se/ to/das/ as/ vis/tas,/ de/ lá/gri/mas...”.

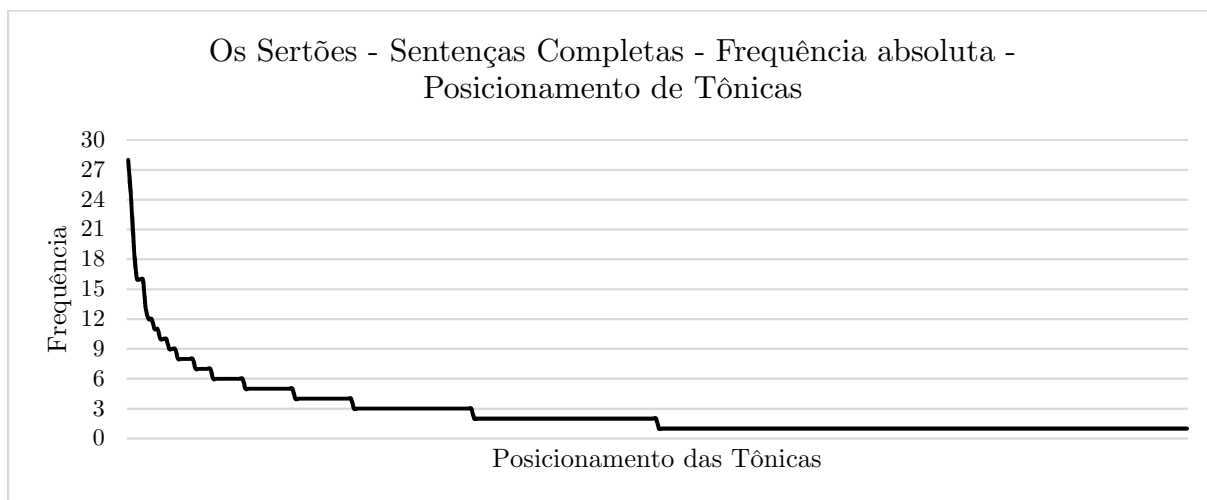


Figura 11: Os Sertões - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas

O gráfico da Figura 11, mostra o comportamento da frequência absoluta de cada posicionamento de tônica ou padrões rítmicos das sentenças com estrutura métrica encontradas pelo MIVES neste experimento. O gráfico mostra que existe uma concentração de alguns padões enquanto outros ocorrem apenas uma vez. Dentre os padrões rítmicos conhecidos da poesia destacam-se os 40 heroicos, 19 martelos, 13 sáficos, 6 gaítas galega e 31 alexandrinos. A Tabela 6 apresenta o ranking dos 10 padrões rítmicos que ocorreram com mais frequência.

Tabela 6: Os Sertões – Sentenças Completas - Ranking dos 10 Posicionamento de Tônicas

Posicionamento de Tônicas	Frequência
2 6 10	28
3 7 11	24
3 6 10	19
3 7 10	16

2 7 10	16
2 7 11	16
4 8 10	13
2 6 8 11	12
1 6 10	12
2 5 10	11

Os resultados obtidos pelo MIVES, nesta etapa da experiência, mostraram as equivalências de metros entre os versos de sentenças completas apresentadas por Campos e o MIVES. Almeida já anunciara em seu artigo, a preferência de Euclides da Cunha por versos deca e dodecassilábicas. Analisando as 652 estruturas de versificação reportadas pelo MIVES, apenas 45 foram escandidas unicamente como hendecassílabos, ou seja, aproximadamente 92,99% das estruturas de versificação identificadas são constituídas por estruturas deca e dodecassilábicas. Os resultados apresentados pelo MIVES vão ao encontro das palavras de Almeida, quanto às preferências de Euclides da Cunha, mostrando dessa forma a eficiência no processo de busca por esse tipo de estrutura executada pelo MIVES.

Etapa 2: Início de Sentença com Estrutura Métrica

No trabalho de [Campos 1996], as estruturas nos inícios de sentenças são escassas. Conforme descrito em nossa Metodologia, ao fazer a busca por estruturas de versificação de início de sentença, o MIVES considera como resultado válido as estruturas de versificação que terminam com os símbolos de pontuação já descritos neste documento. Por esse motivo, de um total de 24 estruturas de versificação apresentadas por Campos, foram considerados 14 trechos iniciais de sentença por estarem delimitados por símbolo de pontuação. Do total de 14 estruturas analisadas, 11 foram escandidas com metros equivalentes aos apresentados por Campos e três com métrica superior a 12 sílabas. A Tabela 7, apresenta as estruturas de versificação apresentadas por Campos em sua primeira coluna, enquanto a segunda e terceira coluna, respectivamente, apresentam a forma escandida e seus respectivos metros apresentados pelo MIVES.

Tabela 7: Relação de Estruturas de Versificação Etapa 2 - Primeiro experimento

Transertões	MIVES	Nº Sílabas Poéticas
O planalto central do Brasil desce,	O/ Pla/ nal /to/ Cen/ tral / do/ Bra/ sil / des /ce,	10
Varada a estreita faixa de cerrados,	Va/ ra /da a es/ trei /ta/ fai /xa/ de/ ce/ rra /dos,	10
	Va/ ra /da a es/ trei /ta/ fa /i/x <i>a</i> / de/ ce/ rra /dos,	11

	Va/ ra /da a es/ tre /i/ta/ fa /i/xa/ de/ ce/ rra /dos,	12
Na postura, no gesto, na palavra,	Na/ pos/ tu /ra,/ no/ ges /to,/ na/ pa/ la /vra,	10
Rebrilham longas noites nas chapadas,	Re/ bri /lham/ lon /gas/ noi /tes/ nas/ cha/ pa /das,	10
	Re/ bri /lham/ lon /gas/ no /i/tes/ nas/ cha/ pa /das,	11
Espécie de grande homem pelo avesso,	Es/ pé /cie/ de/ gran /de ho /mem/ pe /lo a/ ve /sso,	10
	Es/ pé /cie/ de/ gran /de/ ho /mem/ pe /lo a/ ve /sso,	11
Na claridade mortecida dos braseiros,	Na/ cla/ri/ da /de a/mor/te/ ci /da/ dos/ bra/ sei /ros,	12
O tropear soturno das fileiras,	O/ tro/pe/ ar / so/ tur /no/ das/ fi/ lei /ras,	10
	O/ tro/pe/ ar / so/ tur /no/ das/ fi/ le /i/ras,	11
Os demais, sucumbidos de fadigas,	Os/ de/ mais ,/ su/cum/ bi /dos/ de/ fa/ di /gas,	10
O Sol irradiava a pino sobre a terra,	O/ Sol / i/rra/di/ a /va a/ pi /no/ so/bre a/ ter /ra,	12
Transverberando nas rochas expostas,	Trans/ver/be/ ran /do/ nas/ ro /chas/ ex/ pos /tas,	10
Sob a adustão dos dias ardentíssimos,	Sob/ a a/dus/ tão / dos/ di /as/ ar/den/ tí /ssi/mos,	10
	Sob/ a/ a/dus/ tão / dos/ di /as/ ar/den/ tí /ssi/mos,	11

Os versos “Chegavam, estropiados, da jornada longa, ” e “À luz crua dos dias sertanejos, ” escandidos pelo MIVES, respectivamente como “Che/**ga**/vam,/ es/tro/pi/**a**/dos,/ da/ jor/**na**/da/ **lon**/ga,” 13 sílabas e “À/ **luz**/ **crua**/ dos/ **di**/as/ ser/ta/**ne**/jos”, 9 sílabas são exemplos de versos que não foram incluídos entre os resultados reportados ao MIVES por não possuírem os metros desejados para esta etapa do experimento, entre 10 e 12 sílabas poéticas.

Além dos versos apresentados na Tabela 7, neste experimento o MIVES retornou entre

decassílabos e dodecassílabos (estruturas no início das sentenças) 1746 estruturas de versificação. “Urdiam-se estranhos episódios”.(Ur/**di**/am/se/ es/**tra**/nhos/e/pi/**só**/dios.), “As notas das cornetas, convulsivas,”(As/ **no**/tas/ das/ cor/**ne**/tas,/ con/vul/**si**/vas,) são exemplos de decassílabos, enquanto as estruturas de versificação “Fracionados, galgando penhascos a pulso,” (Fra/cio/**na**/dos,/ gal/**gan**/do/pe/**nhas**/cos/ a/ **pul**/so,) e “Equiparavam aos flagelos naturais,” (E/qui/pa/**ra**/vam/aos/ fla/**ge**/los/ na/tu/**rais**,) exemplos de autênticos dodecassílabos apresentados pelo MIVES neste experiência, são exemplos de versos não relacionados por Campos mas que foram identificados pelo MIVES.



Figura 12: Os Sertões – Início de Sentenças - Distribuição da Frequência Absoluta.

Para se ter uma ideia de como estão distribuídos os inícios de sentenças com estrutura métrica ao longo do texto, pode-se analisar o gráfico apresentado na Figura 12. Neste gráfico fica evidente uma concentração de sentenças com estrutura métrica nos intervalos com maior amplitude nos valores de frequência. Embora exista uma maior densidade em alguns intervalos, não existem valores de frequência que possam determinar um ponto de alta concentração destas estruturas, as áreas de maior e menor concentração possuem, respectivamente, um total de 57 e 23 estruturas de versificação, com média das frequências de 39,97 enquanto um desvio padrão 7,33. O gráfico da Figura 13 permite visualizar que existem poucas regiões do texto onde não foram identificadas sentenças com estruturas métricas, apesar de existirem alguns espaços com relativa distância entre as sentenças com estrutura métrica.

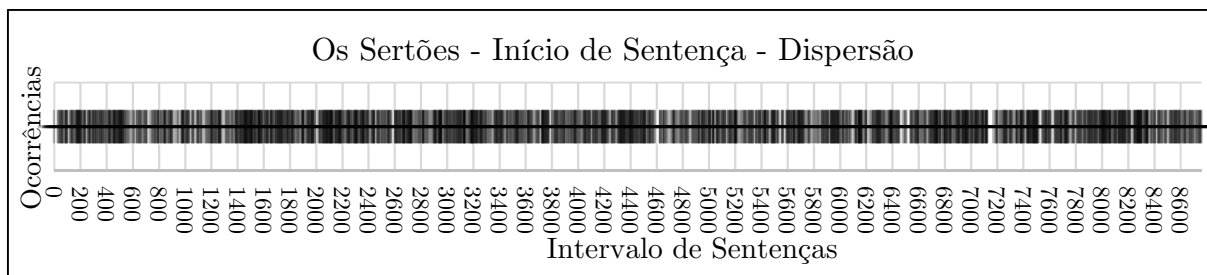


Figura 13: Os Sertões – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.

Além das análises apresentadas até este ponto, através do gráfico apresentado na Figura 14 é possível analisar a frequência do posicionamento das tônicas das estruturas de versificação apresentadas pelo MIVES para os inícios de sentença com estrutura métrica. A Figura 14 apresenta, uma visão geral de como ocorre o comportamento das ocorrências dos Posicionamento de Tônicas a

Tabela 8 apresenta um ranking dos tipos com maior frequência. Dentre outros, dois aspectos podem ser observados neste gráfico, (a) a alternância entre fortes e fracas na maioria dos padrões apresentados e; (b) evidente quantidade de heroicos (105 ocorrências), gaita galega (29 ocorrências), sáficos (22 ocorrências) e alexandrinos (88 ocorrências).

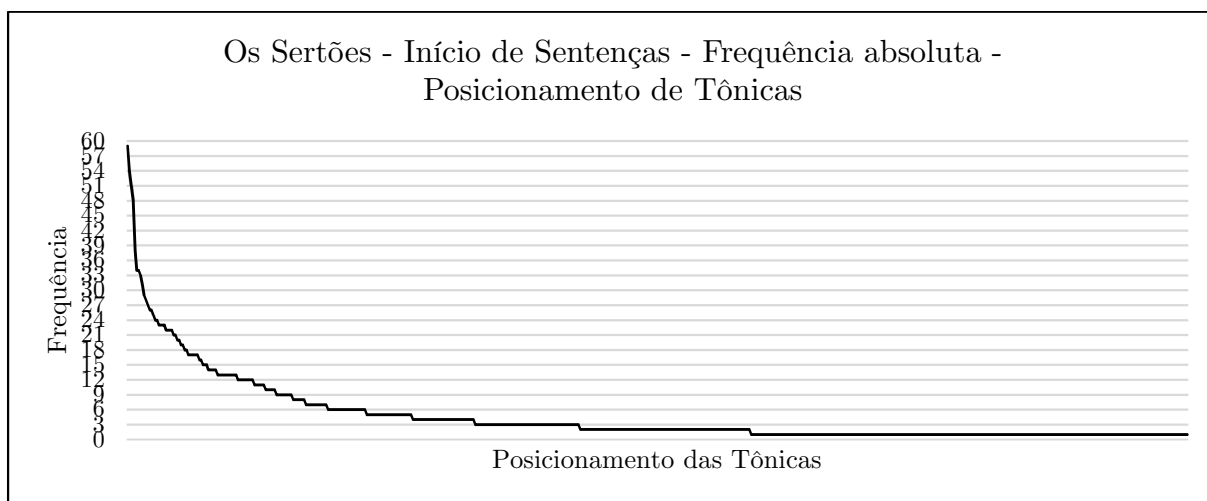


Figura 14: Os Sertões - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.

Tabela 8: Os Sertões – Início de Sentenças - Ranking dos 10 Padrões Rítmicos

Posicionamento de Tônicas	Frequência
3 7 11	59
2 6 10	54
3 6 10	51
3 7 10	48
2 7 11	38
4 7 11	34
3 8 11	34
2 7 10	33
3 6 11	31
4 7 10	29

Além do que já foi apresentado até este ponto, outro fator que merece destaque é a predominância dos decassílabos e dodecassílabos entre os resultados. Das 1746 estruturas de versificação apresentadas como resultado, apenas 129 são unicamente de hendecassílabos, enquanto 1617, referente a cerca de 92,33% foram escandidos como decassílabo ou dodecassílabo. Assim como foi destacado na experiência anterior, esses resultados respaldam a taxa de identificação do MIVES, por conta da preferência de Euclides da Cunha por esses metros conforme destacado por Guilherme de Almeida.

Etapa 3: Finais de sentenças com Estrutura Métrica

Nesta última etapa, foram buscadas pelo software as estruturas de versificação listadas em maior número, tanto por [Almeida 1946] quanto por [Campos 1996]. Este último apresentou um total de 122 estruturas de versificação. Desse total, 14 foram excluídas por conta da quantidade excessiva de sílabas que ultrapassa a métrica convencional. São exemplos destas estruturas, os segmentos escandidos: “e/xu/be/**ran**/do/flo/ra/**ção**/ ri/**den**/te em/ **me**/io/ da/ de/**sor**/dem/ tro/pi/**cal**.”, 20 sílabas; “gan/da/i/e/i/ros/ de/ **to**/dos/ os/ ma/**ti**zes, / re/ci/**di**/vos/ de/ **to**/dos/ os/ de/**li**/tos.”, 23 sílabas; “a i/nu/ma/**ção**/ da/ flo/ra/ mo/ri/**bun**/da, en/te/**rran**/do/se os/ **ca**/u/les/ **pe**/lo/ **so**/lo.”, 21 sílabas, e; “**per**/tur/ba/se em/ a/**ta**/lhos, / on/du/**lan**/te, / tor/ne/**jan**/do /sem/ **nú**/me/ro/ de en/**cos**/tas.”, 21 sílabas.

Das 108 estruturas restantes, consideramos 56 por estarem delimitadas por símbolos de pontuação. Do total de 56 estruturas analisadas, 50 foram escandidas pelo MIVES com metros equivalentes aos apresentados por Campos e apenas 6 com métrica superior a 12 sílabas. A Tabela 9, apresenta alguns exemplos das estruturas de versificação apresentadas por Campos em sua primeira coluna, enquanto a segunda e terceira coluna, respectivamente, apresentam a forma escandida e seus respectivos metros apresentados pelo MIVES.

Tabela 9: Relação de Estruturas de Versificação Etapa 3 - Primeiro experimento

Transertões	MIVES	Nº Sílabas Poéticas
o aspecto atormentado das paisagens.	o as/ pec /to a/tor/men/ ta /do/ das/ pa /i/sa/gens.	10
é um vitorioso jovial e forte.	é/ um/ vi/to/ri/ o /so/ jo/vi/ al / e/ for /te.	12
melopéia plangente dos benditos.	na/ me/lo/ péia / plan/ gen /te/ dos/ ben/ di /tos.	10
	na/ me/lo/ péi /a/ plan/ gen /te/ dos/ ben/ di /tos.	11
feitas a um tempo lares	fei /tas/ a um/ tem /po/ la /res/ e/ re/ du /tos.	10

e redutos.	fe/i/tas/ a um/ tem /po/ la /res/ e/ re/ du /tos.	11
malbaratado a vida em toda a linha.	mal/ba/ra/ ta /do a/ vi /da em/ to /da a/ li /nha.	10
	mal/ba/ra/ ta /do/ a/ vi /da em/ to /da a/ li /nha.	11
	mal/ba/ra/ ta /do/ a/ vi /da/ em/ to /da a/ li /nha.	12
seguindo a extensa fila de infelizes...	se/ guin /do a ex/ ten /sa/ fi /la/ de in/fe/ li /zes...	10
	se/ guin /do/ a ex/ ten /sa/ fi /la/ de in/fe/ li /zes...	11
	se/ guin /do/ a/ ex/ ten /sa/ fi /la/ de in/fe/ li /zes...	12
um desmedido semicírculo de assédio.	um/ des/me/ di /do/ se/mi/ cír /cu/lo/ de a/ ssé /dio.	12
como uma nódoa amplíssima de sangue.	co /mo u /ma/ nó /doa am/ plí /ssi/ma,/ de/ san /gue...	10
	co /mo u /ma/ nó /do/a am/ plí /ssi/ma,/ de/ san /gue...	11
	co /mo/ u /ma/ nó /doa am/ plí /ssi/ma,/ de/ san /gue...	12
caiu durante o dia sobre a tropa...	ca/ i /u/ du/ ran /te o/ dia / so /bre a/ tro /pa...	10

Analizamos como estão distribuídas as estruturas encontradas pelo MIVES nesta etapa do experimento. O gráfico da Figura 15, mostra que, neste caso, existe uma densidade maior de estruturas em alguns intervalos onde os valores de frequência sofrem amplitude. Vale ressaltar que, embora existam áreas de concentração de sentenças com estrutura métrica no texto, a maior diferença de concentração é de 32 estruturas sobre uma janela de 200 sentenças. A maior concentração é de 59 enquanto a menor é de 27 estruturas com uma média aproximada das frequências de 42 e desvio padrão de $\pm 5,69$ regiões do texto que se distanciam da média de distribuição. Isso mostra que, embora não exista uma distribuição uniforme, as estruturas de versificação aparecem com certa regularidade ao longo do texto.

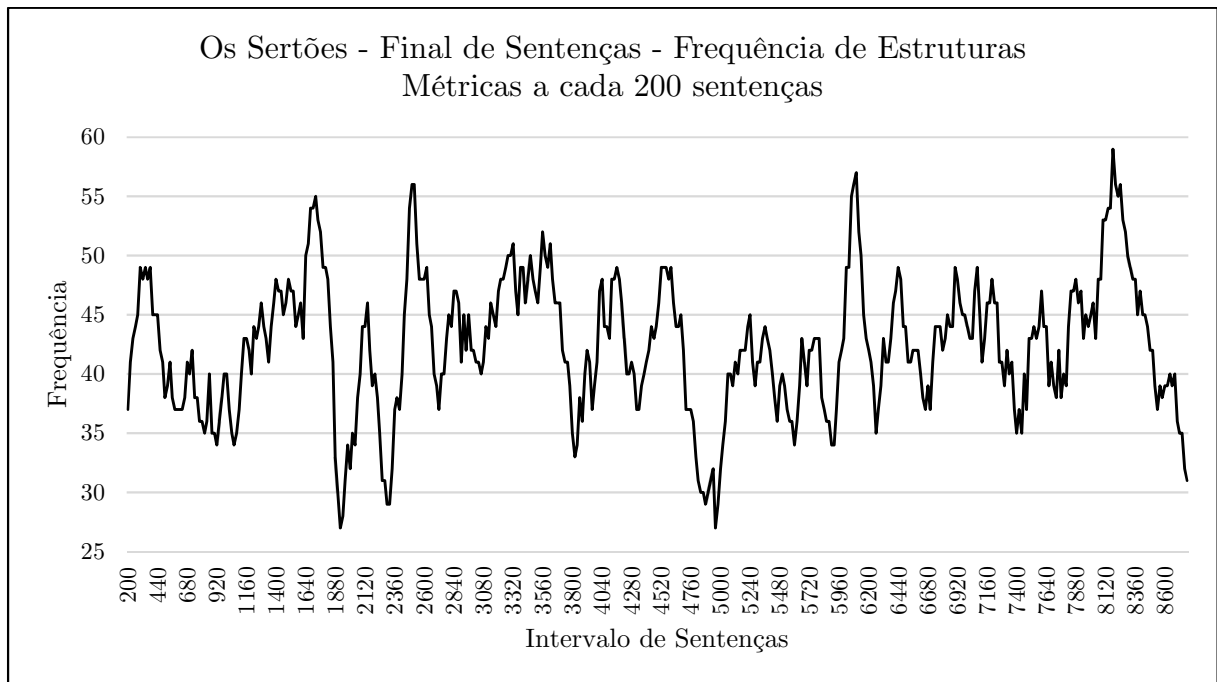


Figura 15: Os Sertões – Final de Sentenças - Distribuição da Frequência Absoluta.

A análise de como estão distribuídas, ao longo do texto, as sentenças com estrutura métrica de final de sentença podem ser realizadas através do gráfico apresentado pela Figura 16. Por conta do número de sentenças desta natureza retornadas pelo MIVES, 1846, e também por conta das limitações impostas pelas dimensões físicas deste documento, não é possível visualizar as regiões do texto em que existem distâncias entre as estruturas métricas reportadas pelo sistema. Para se ter uma ideia de que estas distâncias existem pode-se analisar na Figura 16 as regiões ampliadas do gráfico.

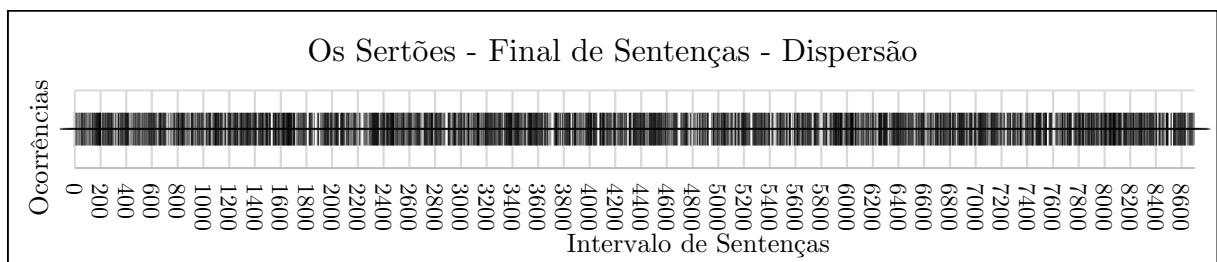


Figura 16: Os Sertões – Final de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.

Assim como nas etapas anteriores deste experimento, analisamos quais os padrões de tônicas se repetem com mais frequência entre as sentenças com estrutura métrica. Também nesta etapa, chama atenção a presença de heroicos, sobretudo aqueles com tônicas nas 3^a, 6^a e 10^a sílabas, um total de 56, seguidos dos martelos, uma variação do heroico, com um total de 55. Incluído essas quantidades de heroico, ao todo o MIVES reportou um total de 125 heroicos. Outro verso decassílabo que aparece com destaque são os sáficos, com 24 ocorrências. Dentre as sentenças com 12 sílabas poéticas destacam-se a presença dos Alexandrinos com 64 ocorrências. No gráfico da Figura 17 é possível analisar qual comportamento geral das frequências absolutas dos

Posicionamento de Tônicas encontradas sobre os finais de sentença com estrutura métrica. A Tabela 10 apresenta um ranking dos 10 Posicionamento de Tônicas mais frequente encontrados neste experimento.

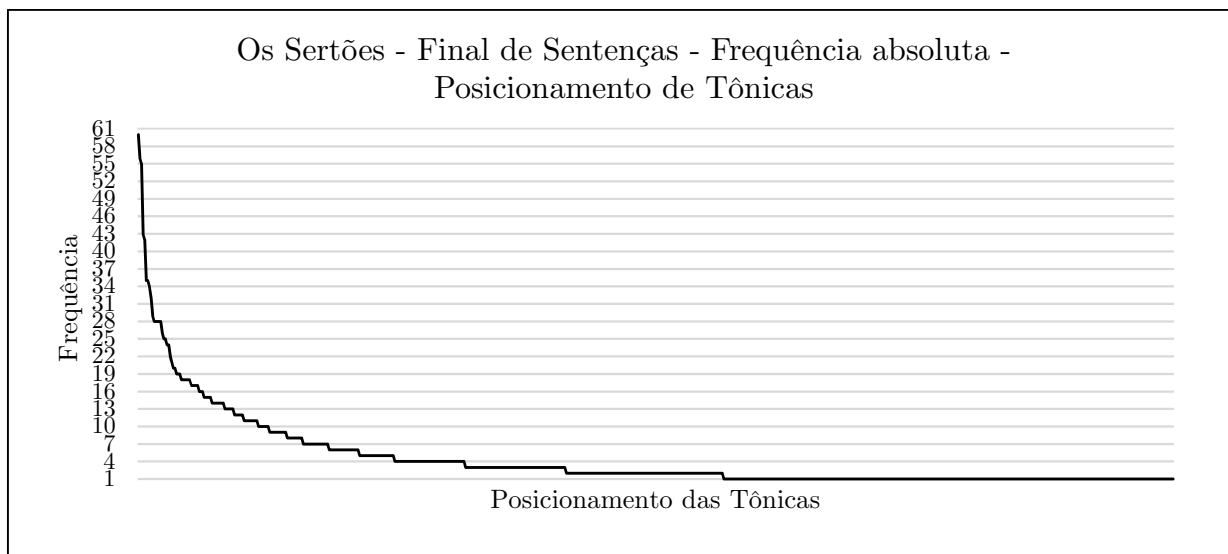


Figura 17: Os Sertões - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.

Tabela 10: Os Sertões – Final de Sentenças - Ranking dos 10 Posicionamento de Tônicas

Posicionamento de Tônicas	Frequência
3 7 11	60
2 6 10	56
3 6 10	55
4 7 11	43
2 7 11	42
4 8 12	35
3 8 11	35
3 7 10	34
3 7 12	32
4 8 11	29

Para concluir, nas 1846 estruturas de versificação apresentadas pelo MIVES, neste experimento, apenas 118 foram escandidas unicamente como hendecassílabos. Enfatizamos mais uma vez a quantidade de estruturas de versificação deca e dodecassilábicas, aproximadamente 94%, um total de 1728 ocorrências.

Conclusão deste Experimento

O objetivo dos experimentos até aqui realizados foi o de validar o processo de busca por estruturas versificação na prosa de Os Sertões utilizando como critério de validação a equivalência de resultados do processo de identificação realizado pelo MIVES com os

versos anunciados nos trabalhos de Almeida[1946] e Campos[1996]. A descrição das etapas apresentadas até este ponto desta dissertação mostrou a eficiência do MIVES no processo.

Além de apresentar resultados compatíveis ao trabalho manual realizado anteriormente, apresentamos outras informações que evidenciam e reforçam a presença de estruturas de versificação em *Os Sertões*, como a grande concentração, constatadas nas três etapas do experimento, de versos heroicos, sáficos e alexandrinos. Considerando todos os resultados apresentados até o momento, a predominância de versos decassílabos e dodecassílabos é outro fator que deve ser evidenciado, pois atesta a eficiência do MIVES, já que, o próprio Guilherme de Almeida destaca em seu trabalho a preferência de Euclides por estes metros. Pelas características das estruturas de versificação apresentadas pelo MIVES estarem de acordo com estudos realizados por Almeida [1946] e por Campos [1996], estes resultados atestam o método utilizado pelo MIVES.

4.3 Análise de Prosa de Euclides

Sabendo da preferência, anunciada por [Campos 1996], de Euclides pelo verso decassilábico e dodecassilábico, repetimos os procedimentos experimentais realizados em *Os Sertões* utilizando os mesmos parâmetros de escansão utilizados na experiência anterior, em duas outras obras de Euclides: “*Contrastes e Confrontos*” e “*À Margem da História*”, publicadas, respectivamente em 1907 e 1909. A motivação para realizar estes experimentos é de verificar se o comportamento versificatório de Euclides, encontrado em *Os Sertões* se repetia em outras obras do autor. “*Contrastes e Confrontos*” reúne artigos publicados originalmente pela imprensa. Já o livro “*À Margem da História*” reúne elementos de pesquisa histórica feita por Euclides, em alguns de seus ensaios. Não foram encontrados, para estas duas obras, estudos da mesma natureza como aqueles realizados por [Almeida 1946] ou [Campos 1996] em *Os Sertões*. Por esse motivo, os resultados apresentados a seguir são aqueles gerados apenas pelo MIVES.

4.3.1 Contraste e Confrontos

Repetimos os mesmos passos da análise feita em *Os Sertões* em *Contraste e Contornos*. Nas três etapas executadas, predominantemente estruturas deca e dodecassilábicas foram encontradas. Para estruturas de versificação formadas por sentenças completas foram encontradas 79, para aqueles de início de sentença 282 e para as de final de 235 estruturas de versificação. A Tabela 11 apresenta exemplos de estruturas de versificação encontrados para sentenças completas. Enquanto o gráfico apresentado na Figura 18 mostra como estão distribuídas, ao longo do texto, estas sentenças.

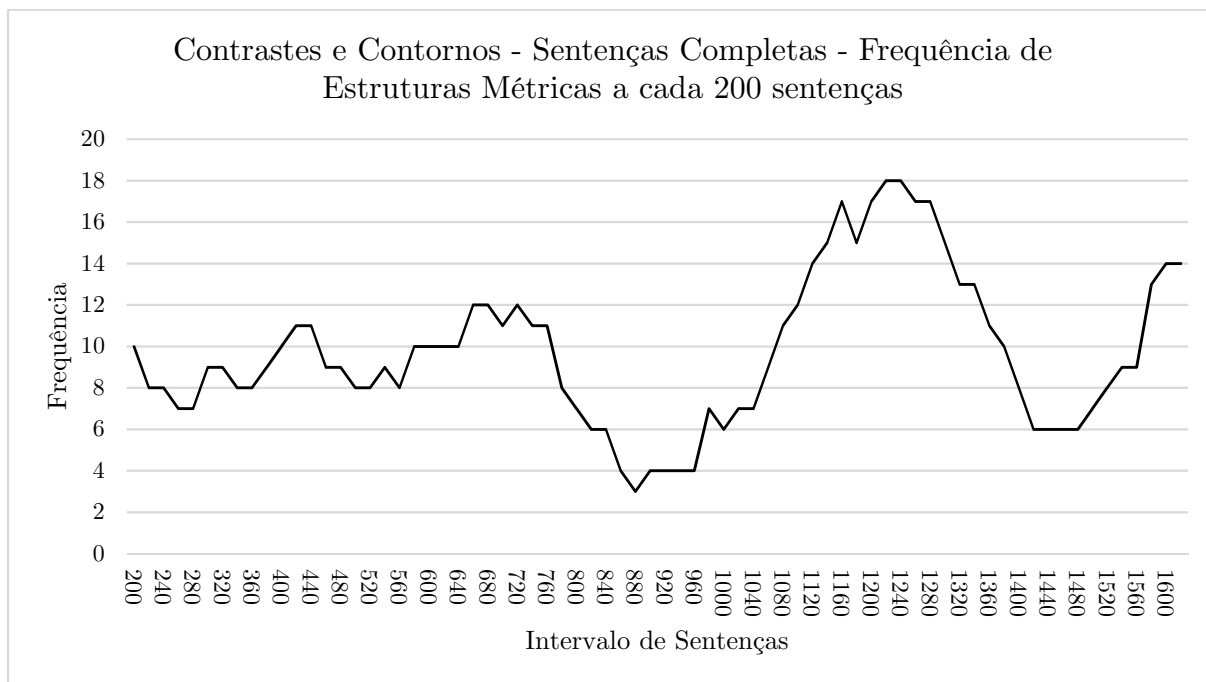


Figura 18: Contrastes e Contornos - Sentenças Completas - Distribuição da Frequência Absoluta.

Os gráficos da Figura 18 e Figura 19 possuem informações complementares a respeito de como estão distribuídas ao longo do texto as estruturas métricas encontradas pelo MIVES. Na Figura 18, por exemplo, temos as frequências absolutas em intervalos de 200 sentenças, da mesma forma que fizemos com os experimentos de Os Sertões. Comparando as informações dos gráficos é possível concluir que existem áreas de maior densidade de sentenças com estrutura métrica e regiões se qualquer ocorrência como uma distância de 110 e 100 sentenças sem qualquer métrica identificada em um total de 1612 sentenças que constituem a obra completa.

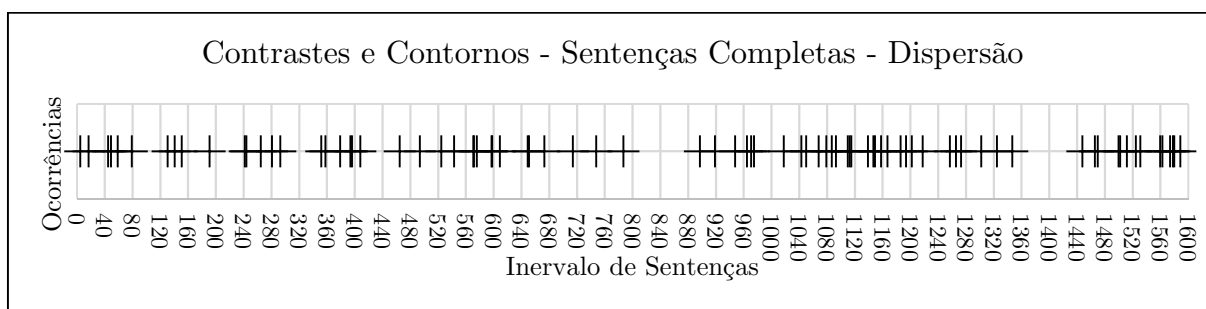


Figura 19: Contrastes e Contornos – Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto.

Tabela 11: Exemplo de Sentenças Completas – Contraste e Contornos

Sentença com Métrica	Completa Estrutura	Escansão segundo MIVES	Nº Sílabas Poéticas
Viu-se, depois, a imponente fragilidade.		Viu- /se,/ de/ pois, / a im/po/ nen /te/ fra/gi/li/ da /de.	12

Arcádia, pátria ideal de toda a gente!	Ar/ cá /dia,/ pá /tria i/de/ al / de/ to /da a/ gen /te!	11
	Ar/ cá /dia,/ pá /tria/ i/de/ al / de/ to /da a/ gen /te!	12
Deflagram na treva os fulgores das descargas.	De/ fla /gram/ na/ tre /va os/ ful/ go /res/ das/ des/ car /gas.	12
Em lances de desespero e audácia	em/ lan /ces/ de/ de/ses/ pe /ro/ e au/ dá /cia.	10
	em/ lan /ces/ de/ de/ses/ pe /ro/ e/ au/ dá /cia.	11
De longe em longe um caminhante.	De/ lon /ge/ em/ lon /ge/ um/ ca/mi/ nhan /te.	10
O deserto invoca o deserto.	O/ de/ ser /to/ in/ vo /ca/ o/ de/ ser /to.	10

Neste experimento foram também verificadas as frequências absolutas do posicionamento das tônicas em Contrates e Contornos. Dentre as classificações mais conhecidas reportadas pelo MIVES estão, 8 heroicos, 1 sáfico e 1 gaita galega e 3 alexandrinos. Um ranking dos 10 Posicionamento de Tônicas com maior frequência pode ser visto na Tabela 12 O gráfico da Figura 20 apresenta uma visão geral da frequência destas estruturas no texto.

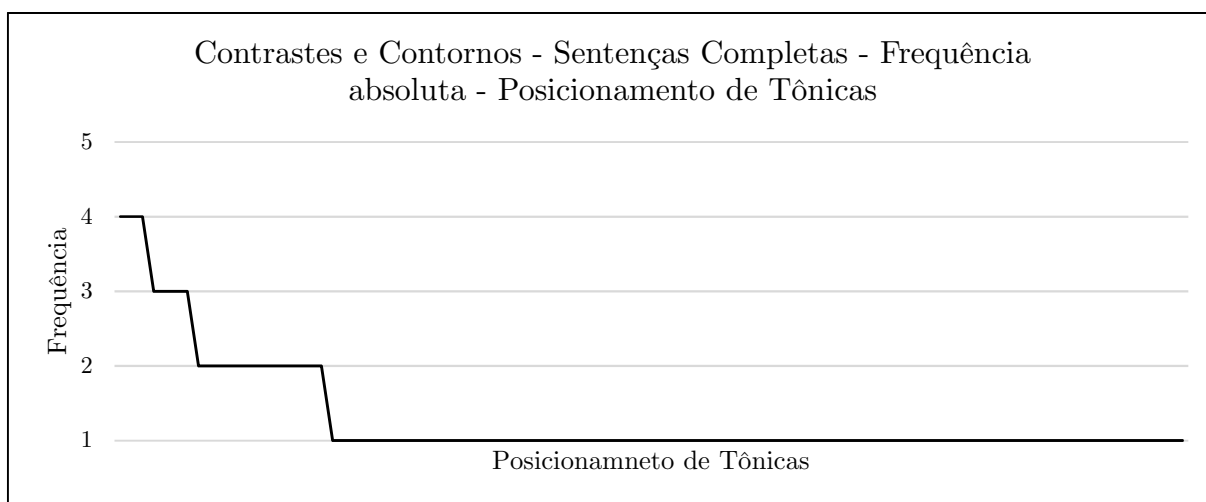


Figura 20: Contrastes e Contornos - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas

Tabela 12: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente –

Posicionamento de Tônicas	Frequência
3 8 11	4
2 6 10	4
3 6 10	4
1 4 8 10	3
4 9	3
4 8 11	3
2 4 8 12	3
3 5 9 11	2
4 10	2
3 7 10	2

Já para as sentenças com estrutura métrica de início de sentença, exemplos são apresentados na Tabela 13. O gráfico da Figura 21 apresenta a destruição de frequência destas estruturas ao longo do texto. Enquanto o gráfico da Figura 22, mostra como estas estruturas estão dispostas ao longo do texto.

Tabela 13: Exemplo de Estruturas de Início de Sentenças – Contraste e Contornos

Início de Sentenças com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Impassível naquele estonteamento	Im/pa/ssí/vel/ na/que/le es/ton/tea/men/to,	10
	Im/pa/ssí/vel/na/que/le es/ton/te/a/men/to,	11
	Im/pa/ssí/vel/ na/que/le/ es/ton/te/a/men/to,	12
Mas, neste concorrer à partilha da terra,	Mas,/nes/te/ con/co/rrer/ à/ par/ti/lha/ da/te/rra,	12
	Bis/mar/ es/bo/ço/u/-se/ sem/ o/ sa/ber,	11
Bismar esboçou-se sem o saber,	Bis/mar/ es/bo/çou/-se/ sem/ o/ sa/ber,	10
E, no momento agudo que vai passando,	E,/ no/ mo/men/to a/gu/do/ que/vai/ pa/ssan/do,	11
	E,/ no/ mo/men/to/ a/gu/do/ que/vai/	12

	pa/ssan/do,	
A atitude entonada, o recacho atrevido,	A a/ti/tu/de en/to/na/da, o/ re/ca/cho a/tre/vi/do,	12
Arcádia, pátria ideal de toda a gente!	Ar/cá/dia,/pá/tria i/de/al/ de/to/da a/gen/te!	11
	Ar/cá/dia,/pá/tria/ i/de/al/ de/to/da a/gen/te!	12

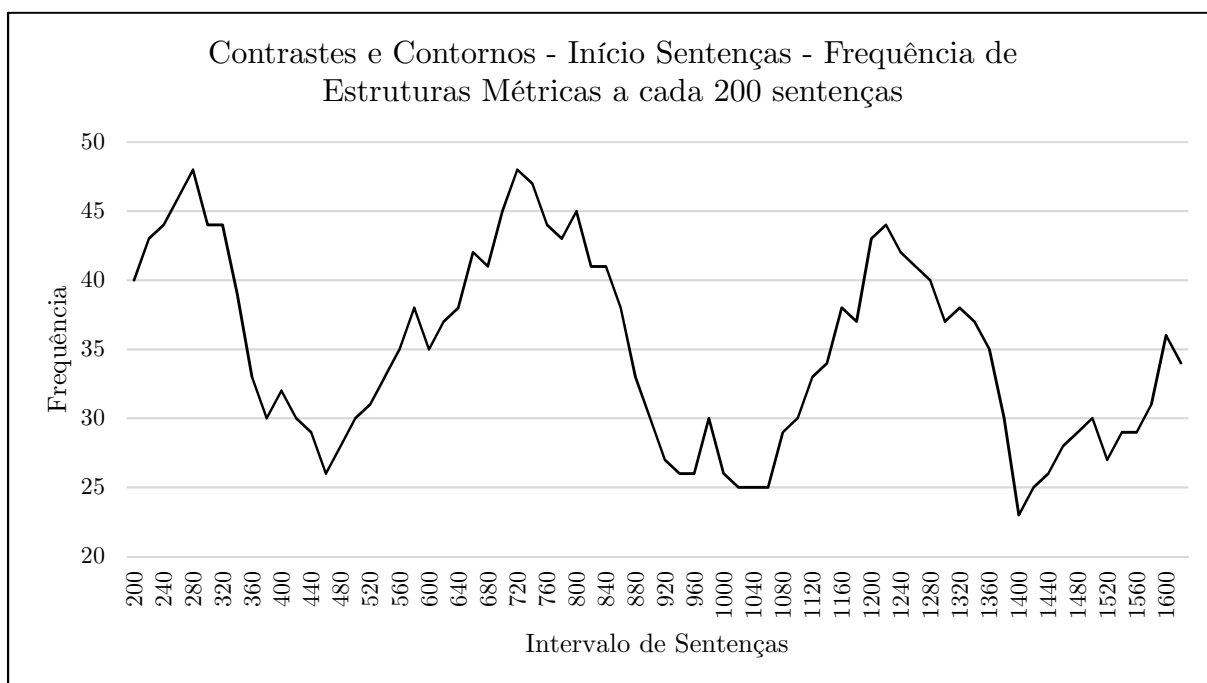


Figura 21: Contrastes e Contornos – Início de Sentenças - Distribuição da Frequência Absoluta.

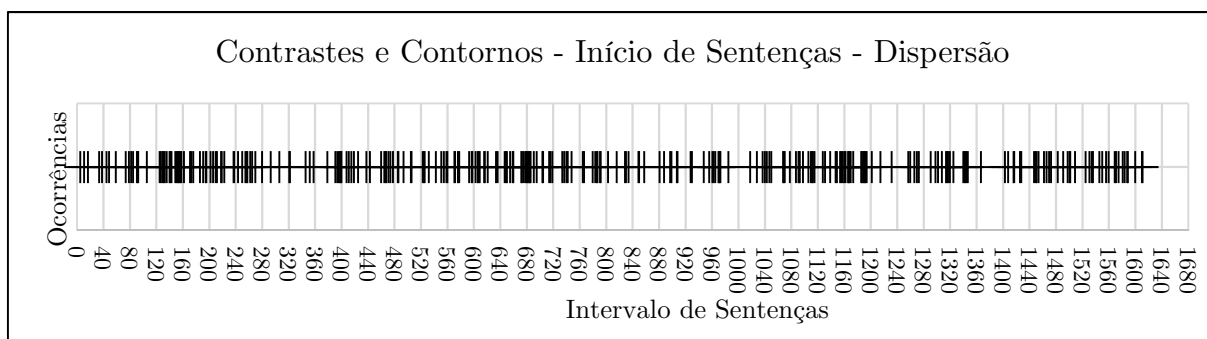


Figura 22: Contrastes e Contornos – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto

Os gráficos Figura 21 e Figura 22 mostram que, embora não aconteçam de maneira uniforme e que existam áreas de maior densidade de estruturas de versificação, estas

estruturas estão distribuídas ao longo texto com intervalos mínimos de 0 e máximo de 48 entre estas estruturas, respectivamente de 0 e 35 sentenças. Mas uma vez, analisamos a frequência absoluta dos tipos de posicionamento de tônicas sobre as sentenças apresentadas pelo MIVES, o resultado pode ser observado pelo gráfico apresentado pela Figura 23.

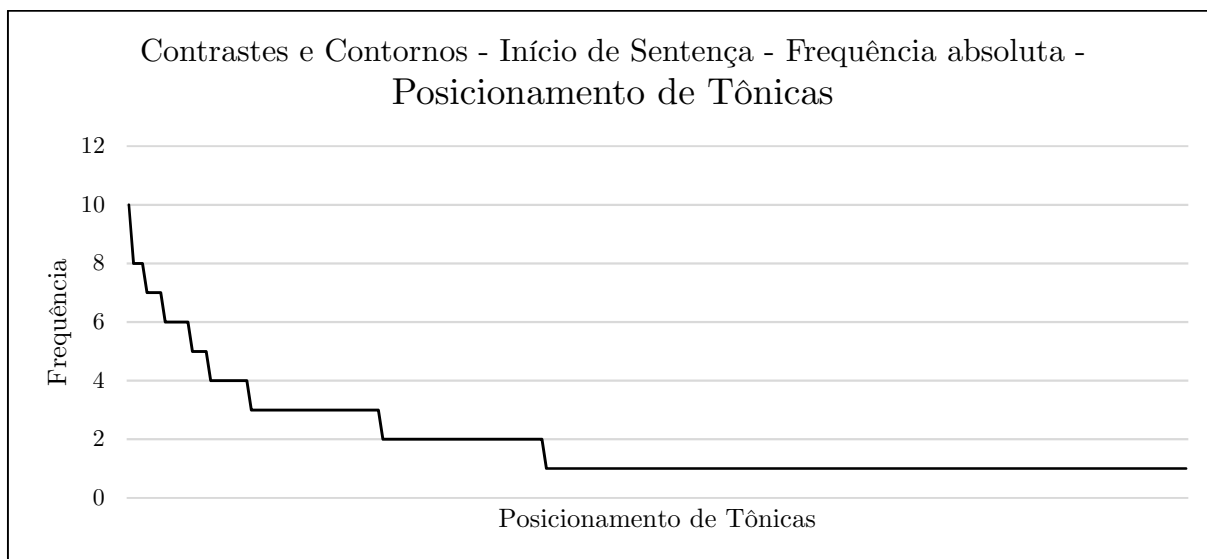


Figura 23: Contraste e Contornos - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.

Quanto a frequência do posicionamento da tônicas entre as estruturas de início de sentença com estrutura métrica destaca-se a presença de 9 heroicos, 8 martelos, 2 sáficos, 6 gaitas galega e 16 alexandrinos. Um ranking dos 10 Posicionamento de Tônicas que ocorreram sobre as sentenas identificadas pode ser visto na Tabela 14.

Tabela 14: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença

Posicionamento de Tônicas	Frequência
4 8 11	10
2 5 10	8
2 6 10	8
3 6 10	8
3 8 11	7
3 5 8 12	7
2 5 8 12	7
4 9 12	7
2 6 11	6
5 8 12	6

Além dos resultados apresentados até este ponto, temos os finais de sentença com estrutura métrica. A Tabela 15 apresenta exemplos de resultados de sentenças desta natureza. Nos gráficos apresentados pelas Figura 24 e Figura 25, é possível visualizar como estas estruturas estão distribuídas ao longo do texto. Na Figura 24 temos a

distribuição de frequências absolutas agrupadas em janelas de 200 sentenças, enquanto do gráfico apresentado pela Figura 25 como estas estão distribuídas ao longo do documento.

Tabela 15: Exemplo de Final de Sentenças – Contrastes e Contornos

Final de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
os traços proeminentes de dois povos.	os/ tra /ços/ proe/mi/ nen /tes/ de/ dois / po /vos.	10
	os/ tra /ços/ proe/mi/ nen /tes/ de/ dois / po /vos.	11
de súbito inclinado para a última.	de/ sú /bi/to in/cli/ na /do/ pa /ra a/ úl /ti/ma.	10
	de/ sú /bi/to/ in/cli/ na /do/ pa /ra a/ úl /ti/ma.	11
	de/ sú /bi/to/ in/cli/ na /do/ pa /ra/ a/ úl /ti/ma.	12
a suspeita receosa das nações.	a/ sus/ pei /ta/ re/ce/ o /sa/ das/ na/ ções .	11
	a/ sus/ pei /ta/ re/ce/ o /sa/ das/ na/ ções .	12
na encruzilhada suspeita de dois mundos...	na en/cru/zi/ lha /da/ sus/ pei /ta/ de/ dois / mun /dos...	
	na en/cru/zi/ lha /da/ sus/ pe /i/ta/ de/ dois / mun /dos...	12
tendo por cima o dilúvio canalizado...	ten /do/ por/ ci /ma o/ di/ lú /vio/ ca/na/li/ za /do...	12
preeminente, a açudada em vasta escala.	pree/mi/ nen /te, a a/çu/ da /da em/ vas /ta es/ ca /la.	10
	pree/mi/ nen /te, a a/çu/ da /da em/ vas /ta es/ ca /la.	11
	pree/mi/ nen /te, a/ a/çu/ da /da em/ vas /ta es/ ca /la.	12

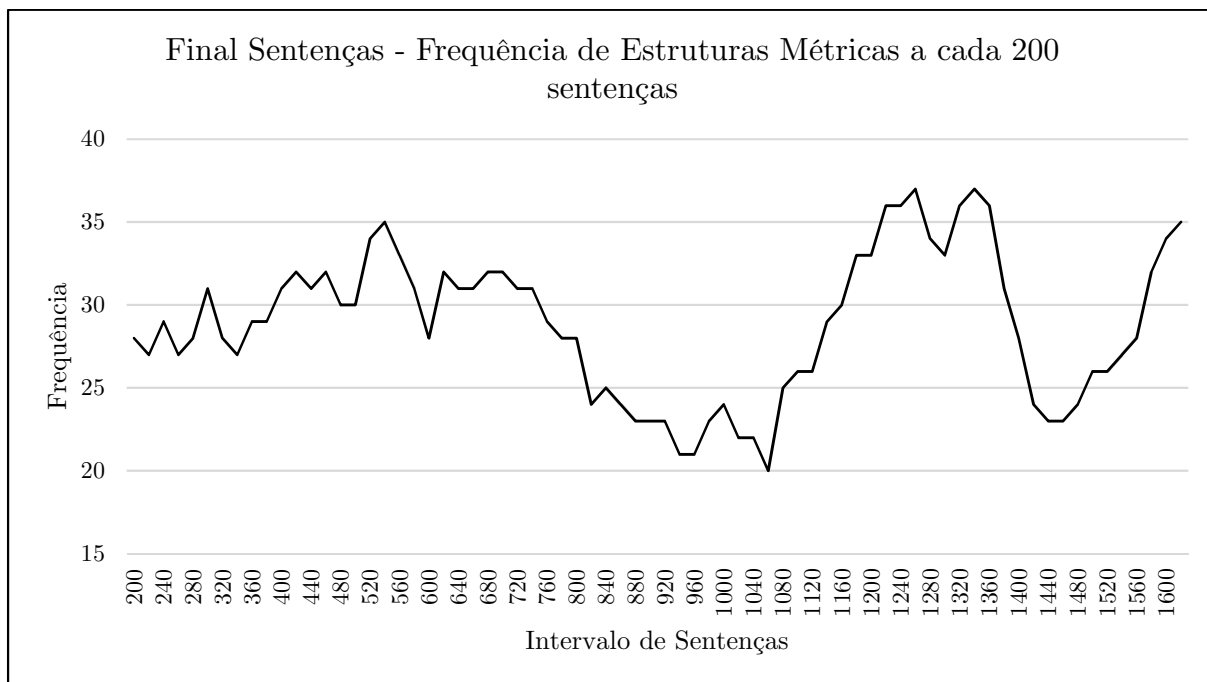


Figura 24: Distribuição de frequências absolutas - Contrastes e Contornos - Final de Sentenças

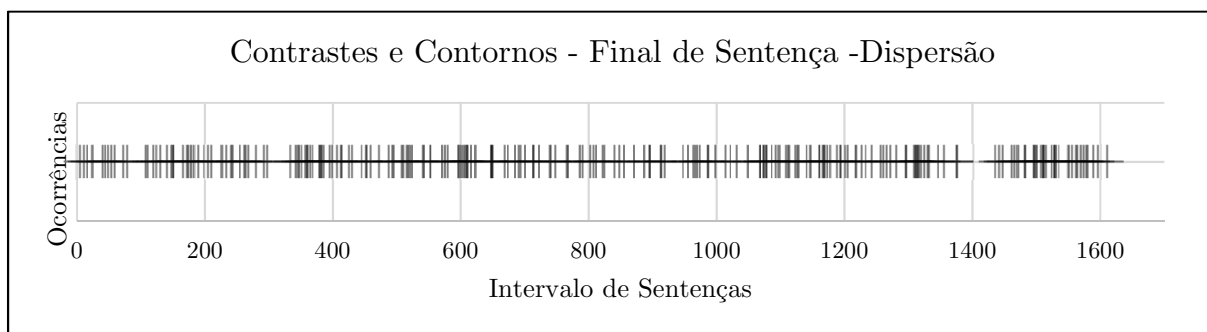


Figura 25: Contrastes e Contornos – Final de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto.

Mas uma vez é possível concluir pelos gráficos da Figura 24 e Figura 25, a existência de várias áreas de concentração de sentenças com estrutura métrica. A maior distância entre sentenças com estrutura métrica é de 59 e a menor 0.

Outro ponto analisado foi a frequência dos padrões rítmicos sobre as estruturas identificadas nos finais de sentença. Dentre os padrões clássicos da poesia foram encontrados 5 heroicos, 5 martelos, 3 saficos, 2 gaíta galega e 7 alexandrinos. Uma visão geral a respeito da frequência dos Posicionamento de Tônicas podem ser vistos no gráfico da Figura 26. A Tabela 16 apresenta um ranking dos 10 padrões que ocorreram com maior frequência.

Comparando os resultados entre os três experimentos sobre Contrastes e Contornos, é possível concluir uma maior concentração sem grandes amplitudes nos espaços de distância entre as sentenças com estrutura métrica que estão nos finais de sentenças.

Tabela 16: Contrastes e Contornos - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença

Posicionamento de Tônicas	Frequência
3 7 11	8
2 5 8 11	6
3 8 11	5
2 6 10	5
3 6 10	5
2 7 10	5
2 7 11	5
2 8 12	5
1 4 8 12	5
3 5 10	5

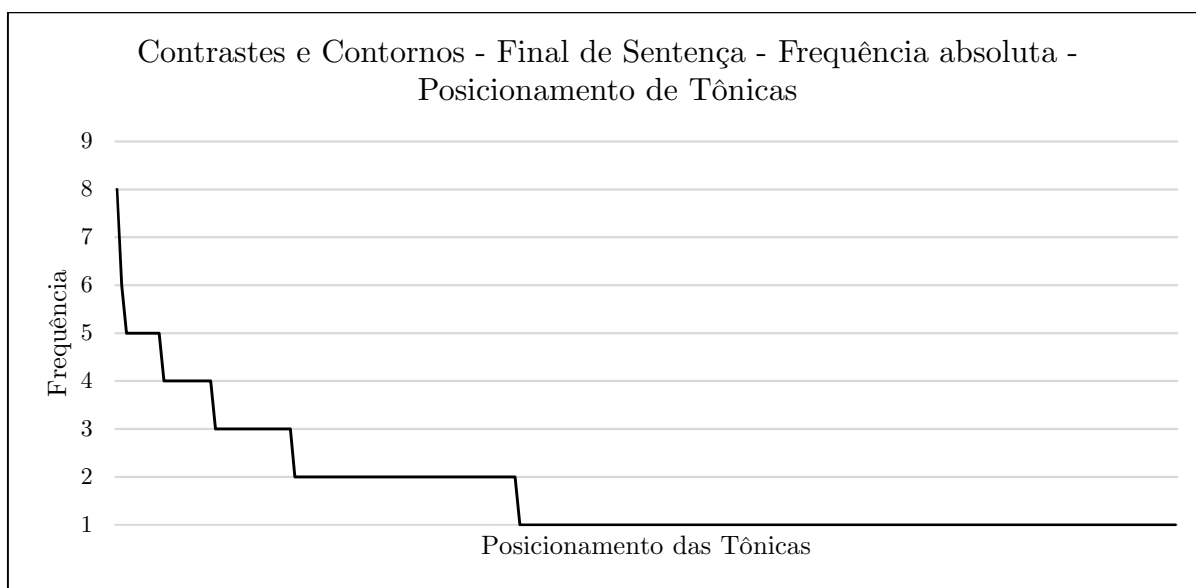


Figura 26: Contrastes e Contornos - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

A Tabela 17 apresenta um resumo, por experiência, do total de sentenças encontradas pelo MIVES com os predominantes percentuais de decassílabos e dodecassílabos:

Tabela 17: Resumo Contrastes e Contornos

Tipo de Busca	Total de Estruturas encontradas	Percentual de decassílabo e dodecassílabo
Sentenças completas	82	95,12%
Início de Sentenças	282	95,04%
Final de Sentenças	235	93,19%

As taxas de identificação obtidas pelo MIVES em Contrates e Contorno nas três etapas ultrapassam os 94,45%. Apesar de não termos parâmetro de comparação, e nos atendo a descrição feita por [Almeida 1946] sobre a preferência de Euclides sobre os versos deca e dodecassilábicos, podemos afirmar que os números aqui apresentados validam mais uma vez o processo de mineração por estruturas de versificação executada pelo MIVES e evidencia mais uma vez o comportamento versificatório de Euclides.

4.3.2 À Margem da História

Seguindo o mesmo critério adotado nas experiências com Os Sertões e Contrastes e Contornos realizamos o experimento em À Margem da História. Neste último experimento, o resultado não foi diferente das experiências anteriores em relação aos metros encontrados. A Tabela 18, Tabela 20 e Tabela 22 apresentam alguns exemplos de estruturas de versificação extraídas de À Margem da História pelo MIVES.

Tabela 18: Exemplo de Sentenças Completas – À Margem da História

Sentença Completas com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
A flora ostenta a mesma imperfeita grandeza."	A/ flo /ra os/ ten /ta a/ mes /ma im/per/ fei /ta/ gran/ de /za.	12
Falo a prosa da minha ciência.	Fa /lo/ a/ pro /sa/ da/ mi /nha/ ci/ ên /cia.	10
As verdades desfecham em hipérboles.	As/ ver/ da /des/ des/ fe /cham/ em/ hi/ pér /bo/les.	10
Revela-no bem um confronto expressivo.	Re/ ve /la-/no/ bem / um/ con/ fron /to ex/pre/ ssi /vo.	11
	Re/ ve /la-/no/ bem / um/ con/ fron /to/ ex/pre/ ssi /vo.	12
Refina a sua astúcia extraordinária.	Re/ fi /na a/ sua / as/ tú /cia ex/tra/or/di/ ná /ria.	11
	Re/ fi /na a/ su /a/ as/ tú /cia ex/tra/or/di/ ná /ria.	12
Aguarda o viajante macabro.	A/ guar /da/ o/ vi/a/ jan /te/ ma/ ca /bro.	10

Na primeira etapa deste experimento um total de 76 sentenças com estrutura métrica, distribuídas ao longo do texto, foram apresentadas entre os resultados. O gráfico apresentado pela Figura 27 mostra a frequência absoluta acomodada em intervalos de 200 sentenças. O gráfico da Figura 28 apresenta como estas estão distribuídas ao longo do texto. No gráfico da Figura 27 é possível visualizar que existem regiões com uma densidade maior de estruturas enquanto outras com menores ocorrências de sentenças com estrutura métrica. Para reforçar essa conclusão basta realizar uma análise no gráfico da Figura 28 onde é possível visualizar intervalos ou janelas de no texto em que as estruturas métricas não estão presentes. O maior intervalo entre as sentenças é de 77 sentenças sem a ocorrência destas com estrutura métrica. Isso mostra que em A Margem da História não existe uma distribuição uniforme destas estruturas.

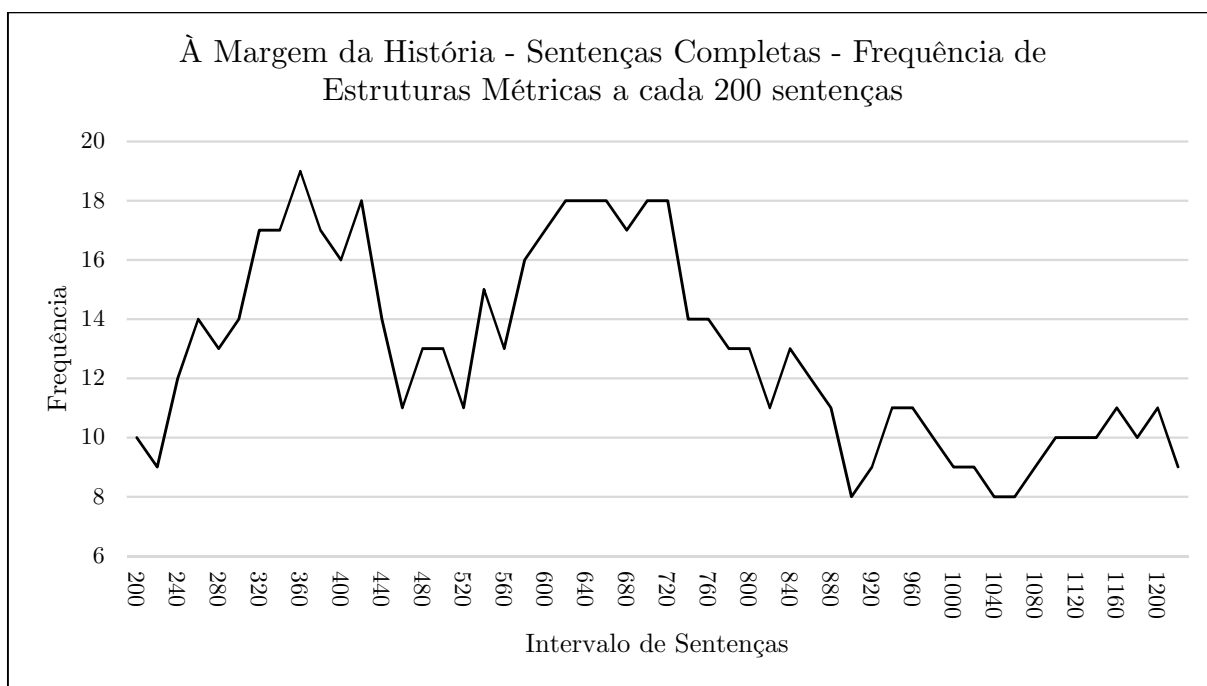


Figura 27: A Margem da História - Sentenças Completas - Distribuição de Frequências Absolutas.

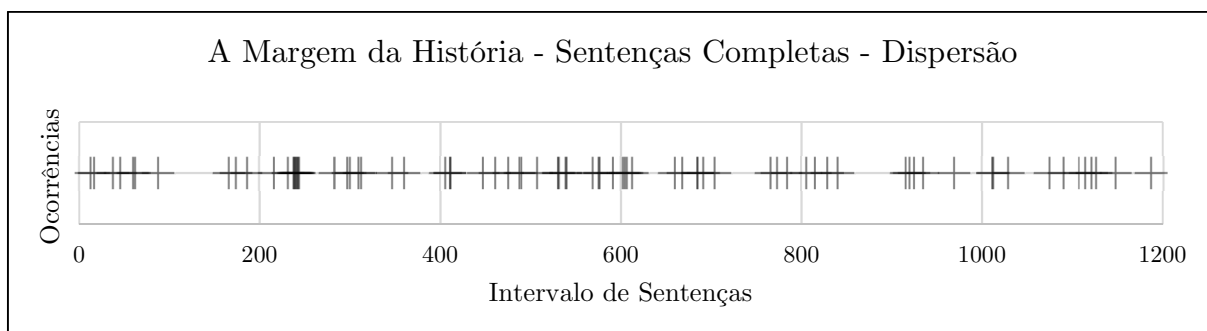


Figura 28: A Margem da História – Sentenças Completas - Distribuição das Sentenças Métricas ao Longo do Texto

Além de avaliar a distribuição das estruturas ao longo do texto, fizemos uma análise para verificar quais os padrões de acentuação tônica aparecem em maior número nesta obra. Não diferente das experiências anteriores, encontramos decassílabos e dodecassílabos de natureza diversa como os 5 heroicos, 1 sáfico e os 5 alexandrinos. O gráfico da Figura 29 apresenta a frequência absoluta da distribuição de tônicas nesta

etapa em que buscamos por sentenças completas com estrutura métrica. A Tabela 19 apresenta um ranking dos 10 posicionamento de Tônicas de maior frequência entre as estruturas identificadas.

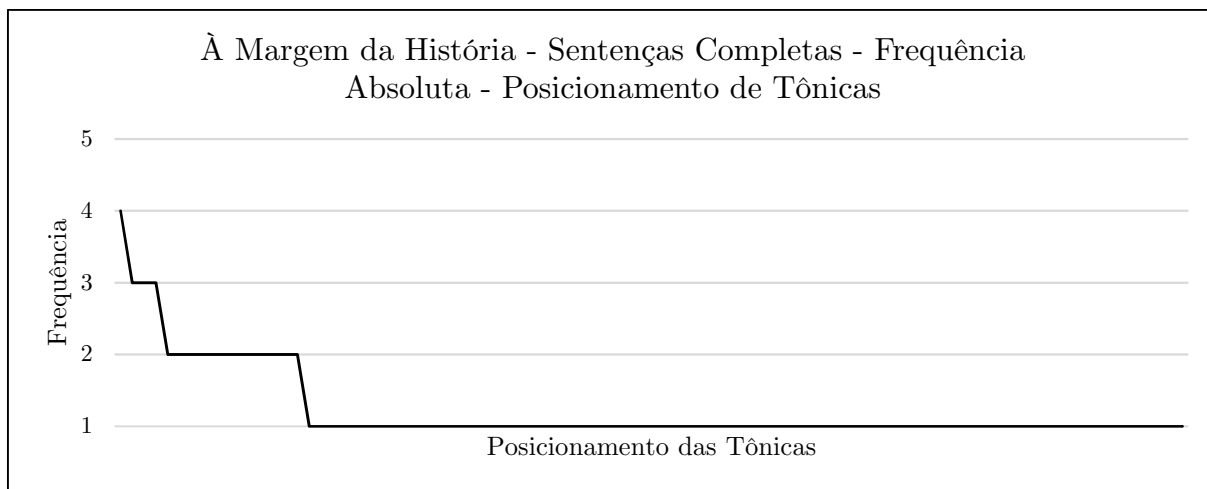


Figura 29: À Margem da História - Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas

Tabela 19: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentença Completas

Posicionamento de Tônicas	Frequência
3 5 10	4
2 6 11	3
2 6 12	3
2 6 10	3
3 7 8 11	2
2 4 7 11	2
3 4 7 10	2
5 9 12	2
2 4 6 11	2
2 4 6 12	2

Tabela 20: Exemplo de Início de Sentenças – À Margem da História

Início com Métrica	Sentença com Estrutura	Escansão segundo MIVES	Nº Sílabas Poéticas
Ao revés da admiração ou do entusiasmo,		Ao/ re/ vés / da ad/mi/ra/ ção / ou/ do en/tu/si/ as /mo,	12
Os cenários, invariáveis no espaço,		Os/ ce/ ná /rios,/ in/va/ri/ á /veis/ no es/ pa /ço,	11

	Os/ ce/ ná /rios,/ in/va/ri/ á /veis/ no es/ pa /ço,	12
Vede esta conta de venda de um homem:	Ve /de es/ta/ con /ta/ de/ ven /da/ de um/ ho /mem:	10
	Ve /de/ es /ta/ con /ta/ de/ ven /da/ de um/ ho /mem:	11
	Ve /de/ es /ta/ con /ta/ de/ ven /da/ de/ um/ ho /mem:	12
A indecisa arqueadura, que logo se forma,	A in/de/ ci /sa ar/quea/ du /ra,/ que/ lo /go/ se/ for /ma,	12
Revela-no bem um confronto expressivo.	Re/ ve /la-/no/ bem / um/ con/ fron /to ex/pre/ ssi /vo.	11
	Re /ve/la/no/ bem / um/ con/ fron /to/ ex/pre/ ssi /vo.	12
Multiplicam-se os casos deste teor,	Mul/ti/ pli /cam-/se os/ ca /sos/ des /te/ te/ or ,	11
	Mul/ti/ pli /cam-/se/ os/ ca /sos/ des /te/ te/ or ,	12

Para a segunda etapa foram encontradas 227 sentenças com estrutura métrica. O gráfico da Figura 30 apresenta como estão distribuídas estas estruturas ao longo do texto. Da mesma forma que nas experiências anteriores, as frequências absolutas dos intervalos apresentados por este gráfico estão organizadas em intervalos ou janelas de 200 sentenças. Além de analisar a frequência absoluta nos intervalos anunciados, é possível visualizar a dispersão das sentenças, com estrutura métrica de início de sentença, ao longo do texto através do gráfico da Figura 31. Neste gráfico também é possível visualizar que existem áreas de concentração de sentenças com estrutura métrica e regiões em que o fenômeno não acontece, com um intervalo de máximo de 27 e uma média 4 sentenças sem a ocorrência de sentenças com estrutura métrica.

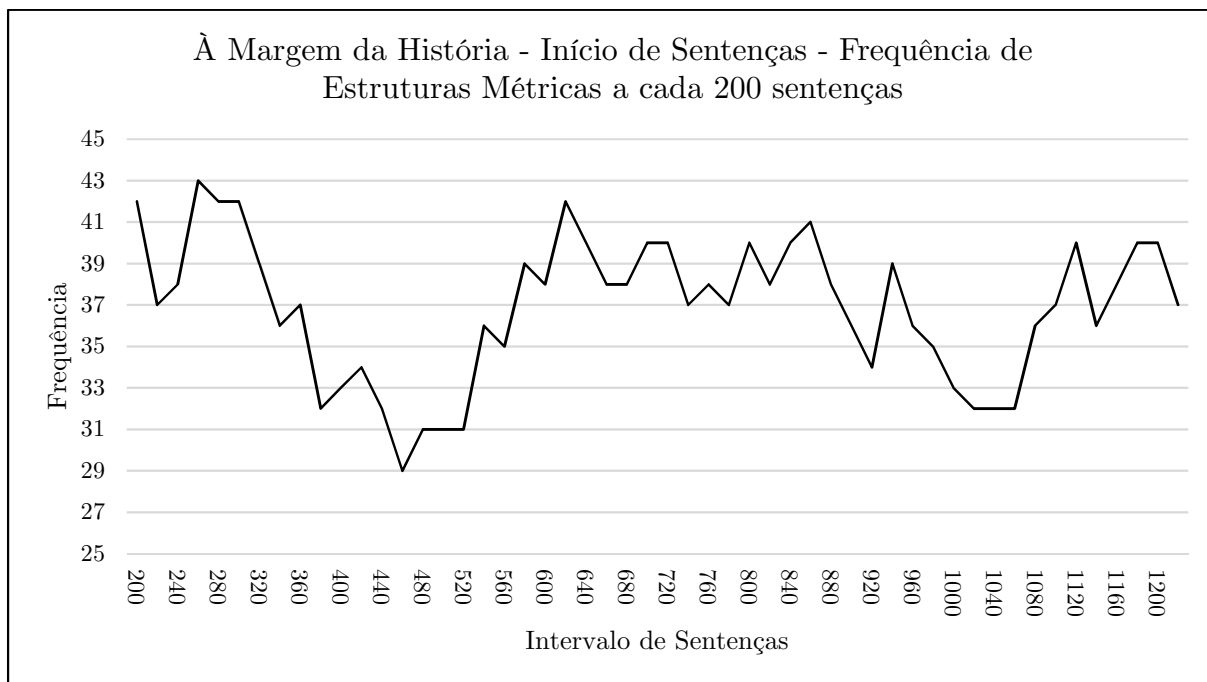


Figura 30: À Margem da História – Início de Sentenças - Distribuição da Frequências Absolutas.

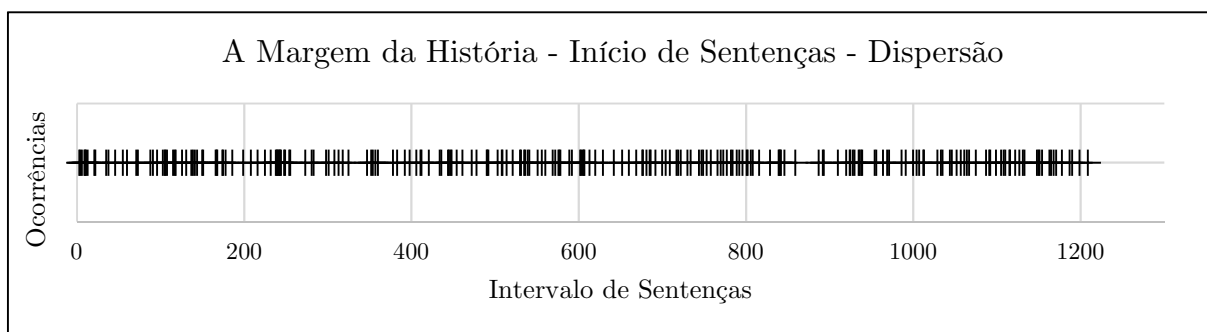


Figura 31: A Margem da História – Início de Sentenças - Distribuição das Sentenças Métricas ao Longo do Texto

Também identificamos nesta segunda etapa do experimento com a À Margem da História formações de tônicas características do verso português, como mais destaque para os 12 heroicos, 6 martelos, 3 sáficos, 3 gaitas galega e cerca de 19 alexandrinos. O gráfico da Figura 32, apresenta o comportamento relacionado a frequência com que os Posicionamento de Tônicas ocorrem entre os resultados, enquanto a Tabela 21 um ranking dos 10 padrões de maior frequência. A Tabela 22 apresenta alguns exemplos de sentenças com estrutura métrica encontradas nesta etapa do experimento.

Tabela 21: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença

Posicionamento de Tônicas	Frequência
2 7 11	9
2 6 12	8
2 6 10	8
2 6 11	7
2 5 8 12	7
4 7 11	7

3 6 10	6
3 7 11	5
2 5 8 11	5
1 4 6 10	5

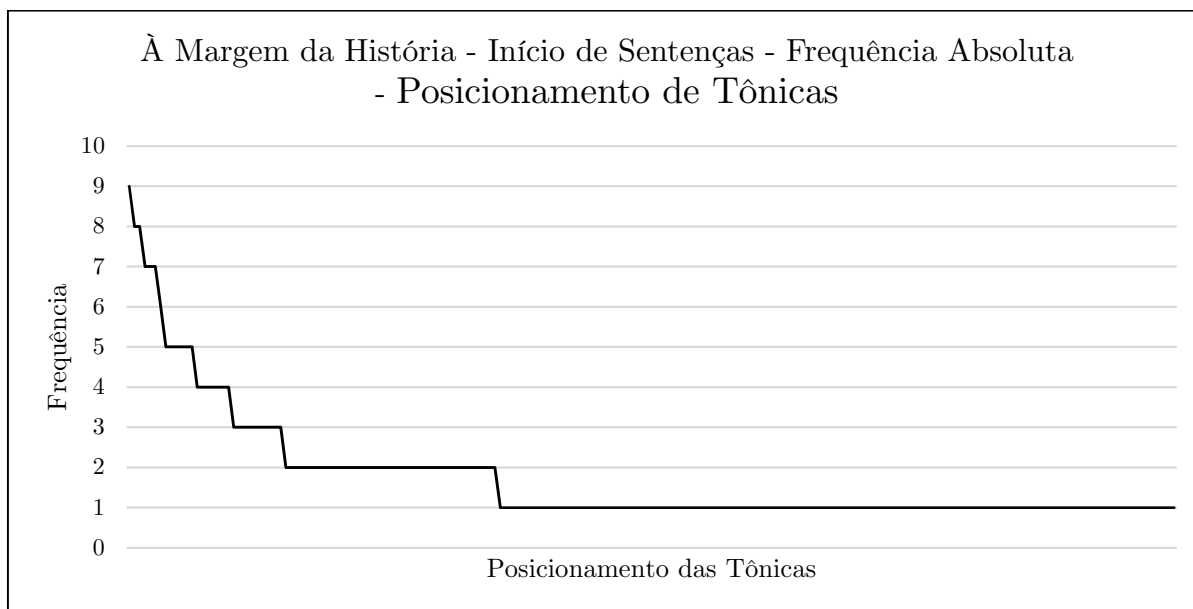


Figura 32: À Margem da História - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas.

Tabela 22: Exemplo Final de Sentenças – À Margem da História

Final com Métrica	Sentença Estrutura	Escansão segundo MIVES	Nº Sílabas Poéticas
	ali, é ainda um intruso impertinente.	a/ li ,/ é a/ in /da um/ in/ tru /so im/per/ti/ nen /te.	11
		a/ li ,/ é/ a/ in /da um/ in/ tru /so im/per/ti/ nen /te.	12
	o que é ainda uma impressão paleozóica.	o/ que é a/ in /dau /ma im/pre/ ssão / pa/leo/ zô /i/ca.	10
		o/ que é a/ in /dau /ma im/pre/ ssão / pa/leo/ zô /i/ca.	11
		o/ que/ é/ a/ in /dau /ma im/pre/ ssão / pa/leo/ zói /ca.	12

e encarrilhou-se nas deduções rigorosas.	e en/ca/rri/ lh ou-/se/ nas/ de/du/ ç ões/ ri/go/ ro /sas.	12
esparsas, a esmo, na amplidão de uma baía.	es/ par /sas,/ a es/mo,/ na am/pli/ d ão/deu /ma/ ba/ i /a.	12
as largas superfícies que atravessa.	as/ lar /gas/ su/per/ f i/cies/ que a/tra/ ve /ssa.	10
	as/ lar /gas/ su/per/ f i/cies/ que/ a/tra/ ve /ssa.	11
todo esse delta lateral se abata.	to /do e/sse/ del /ta/ la/te/ ral / se a/ ba /ta.	10
	to /do/e/sse/ del /ta/ la/te/ ral / se a/ ba /ta.	11
	to /do/e/sse/ del /ta/ la/te/ ral / se/ a/ ba /ta.	12

Para finalizar os experimentos com A Margem da História, foi realizada a busca por sentenças com estrutura métrica no final. Nesta última etapa foram apresentadas pelo MIVES um total de 219 estruturas com sentença métrica, distribuídas ao longo do texto conforme gráfico da Figura 33 e Figura 34. Aqui também é possível concluir que, não existem uma distribuição uniforme destas sentenças ao longo do texto, o que existe são regiões com maiores e menores densidades, com uma distância máxima e mínimas, respectivamente de 27 e 0 sentenças sem estrutura métrica.

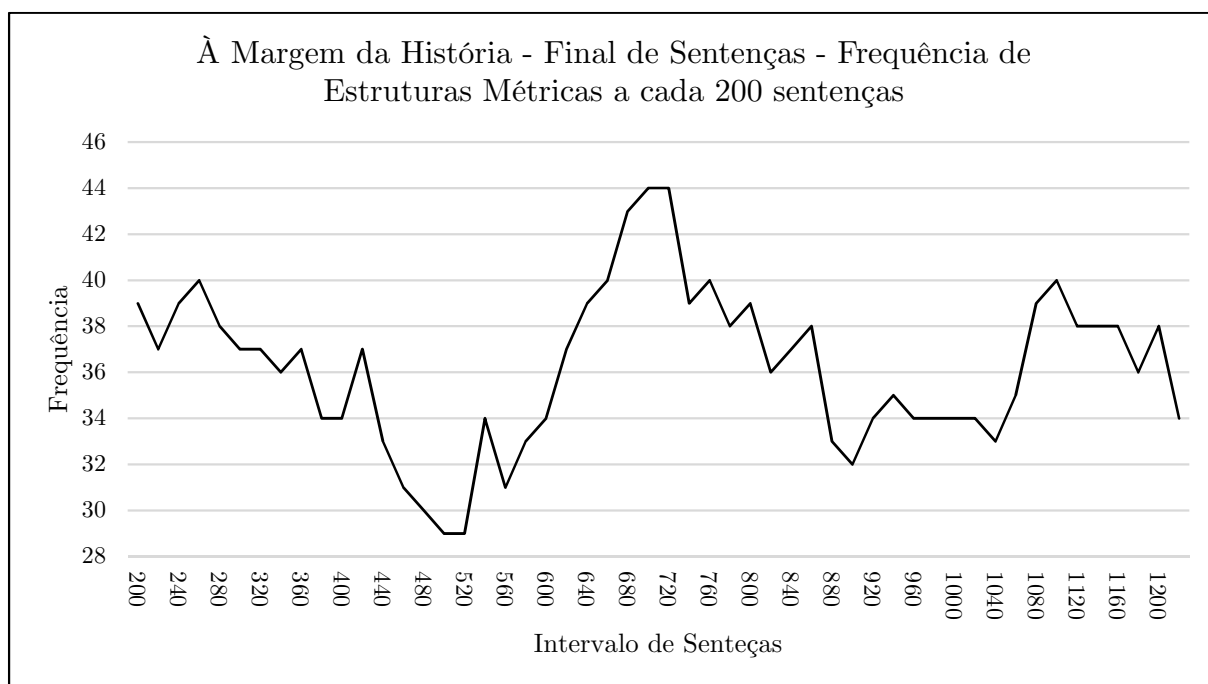


Figura 33: À Margem da História – Final de Sentenças - Distribuição de Frequências Absolutas.

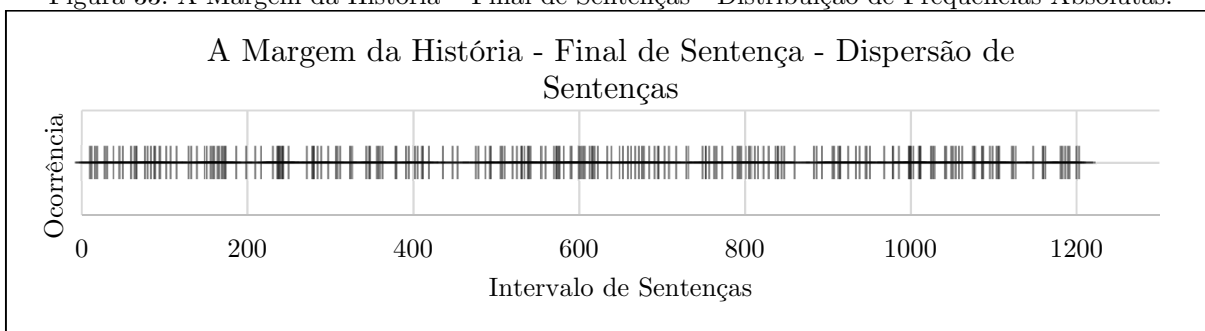


Figura 34: A Margem da História – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto.

Para completar esta etapa verificamos quais eram os padrões de posicionamento de tônicas clássicos da poesia que ocorreram com maior frequência. Neste caso o MIVES apresentou 14 heroicos, 7 martelos, 1 sáfico e 13 alexandrinos. O gráfico da Figura 35 mostra o comportamento da frequência absoluta das formações de tônicas no texto. A Tabela 23 apresenta um ranking 10 padrões de rítmicos de maior frequência.

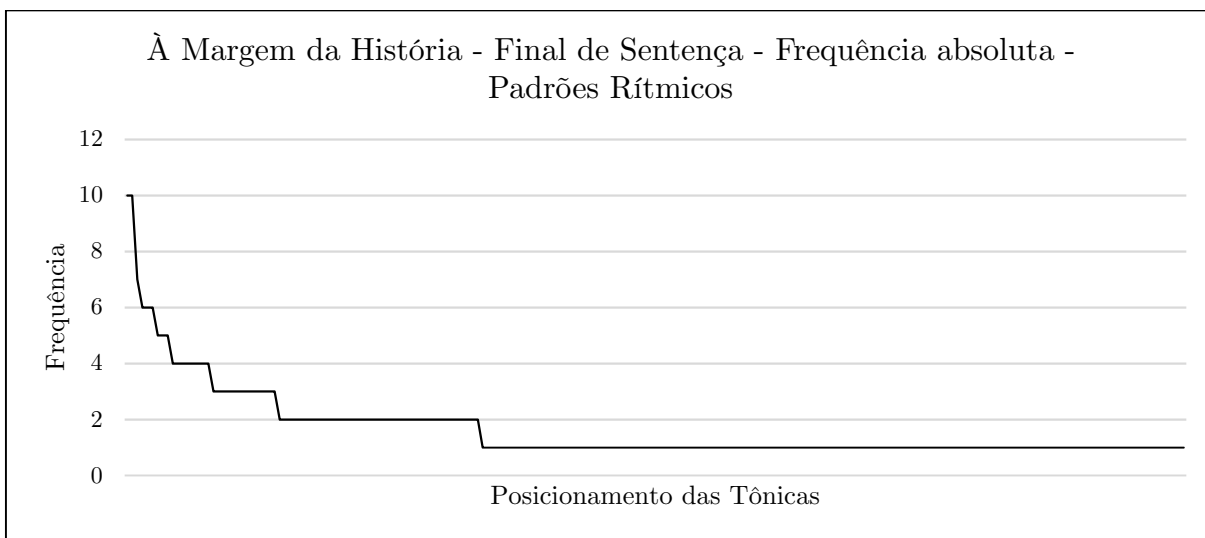


Figura 35: À Margem da História - Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Tabela 23: À Margem da História - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença

Posicionamento de Tônicas	Frequência
2 6 11	10
2 6 10	10
3 6 10	7
3 7 11	6
2 5 8 12	6
2 7 11	6
3 8 12	5
2 6 12	5
2 7 10	5
3 8 11	4

A Tabela 24 apresenta um resumo, por etapa, do total de sentenças encontradas pelo MIVES. Mais uma vez, com os predominantes percentuais de decassílabos e dodecassílabos:

Tabela 24: Resumo À Margem da História

Tipo de Busca	Total de Estruturas encontradas	Percentual de decassílabo e dodecassílabo
Sentenças completas	76	94,74%
Início de Sentenças	227	92,95%
Final de Sentenças	219	92,69%

Com uma média de 93,46% em sua taxa de identificação entre versos decassílabos e dodecassílabos, o MIVES obteve uma taxa de identificação relevante de estruturas de versificação que esperávamos encontrar em mais essa obra de Euclides Cunha.

4.3.3 Comparativo entre as obras de Euclides

Um resumo dos resultados obtidos pode ser visto na Tabela 26. Esta tabela apresenta informações sobre o processamento de Os Sertões e os compara com Contrastes e Contornos e A Margem da História, todas obras de Euclides da Cunha. Em geral, o comportamento versificatório encontrado em Os Sertões, embora não se tenha repetido da mesma forma, esteve presente em todas as suas obras. A frequência das sentenças com estrutura métrica relativa ao número total de sentenças analisadas, entre as três obras, apresentou pequenas diferenças alcançando no máximo 6,1%, conforme pode ser observado no gráfico da Figura 36.

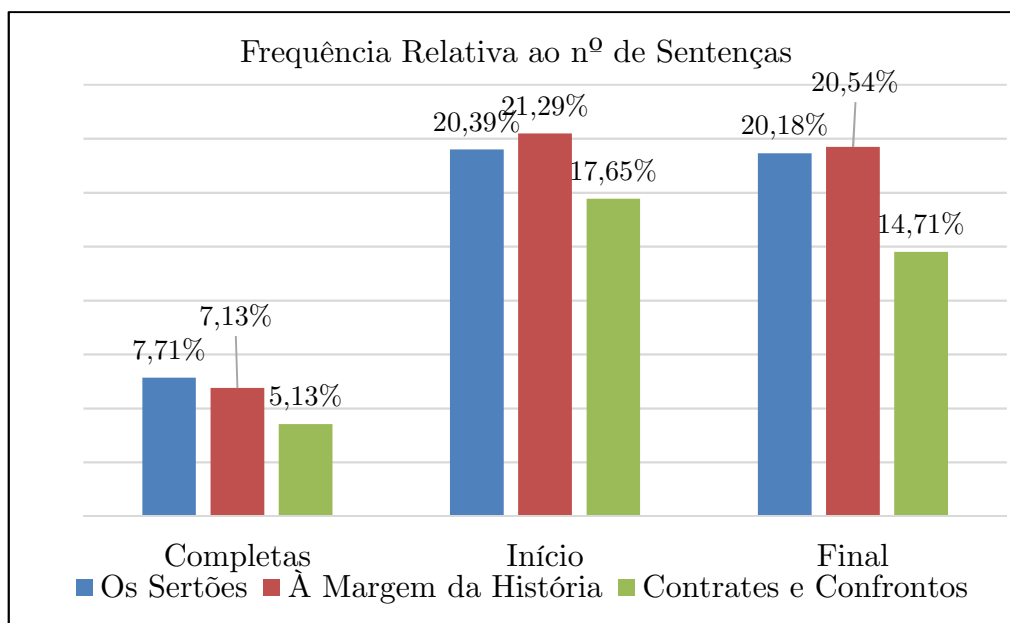


Figura 36: Comparativo entre três Obras de Euclides - Frequência relativa ao total de sentenças.

No caso das sentenças completas, a diferença entre as frequências relativas ao total de sentenças das três obras não ultrapassa 2,97% e para início de sentenças 3,64%. Apenas para os finais de sentenças é possível concluir que há uma distância mais acentuada entre suas frequências, cerca 6,1%, embora a diferença entre Os Sertões e Á Margem da História seja de apenas 2,8%. Além disso, os resultados apresentados na Tabela 26 evidenciam a predominância de estruturas deca e dodecassilábicas, em relação ao total das sentenças encontradas nas obras analisadas, que ultrapassam 90% do total de sentenças métricas apresentadas pelo MIVES.

Além dos comparativos apresentados até este ponto, embora não seja o foco principal deste estudo, vale ressaltar que o MIVES foi capaz de reportar diferentes padrões rítmicos que se repetiram por todas as obras analisadas de Euclides. A Tabela 25 apresenta um resumo dos principais padrões rítmicos encontrados nos experimentos.

Tabela 25: Padrões rítmicos das obras analisadas de Euclides

Obra	Heroico	Martelo	Sáfico	Gaita Galega	Alexandrino
Os Sertões	182	125	59	50	220
Contraste e Contornos	18	17	6	9	32
A Margem da História	30	14	3	4	29

Os valores de padrões rítmicos extraídos de Os Sertões são expressivamente maiores que aqueles extraídos das outras obras. Porém é importante destacar que as obras possuem extensões diferentes, conforme pode ser observado na Tabela 26, enquanto em Os Sertões foram processadas 8564 sentenças em Contrastes e Contorno foram processadas 1598 e em A Margem da História 1066. Se analisarmos a frequência dos padrões rítmicos conhecidos em relação ao número de sentenças processadas em cada obra, será possível constatar que a frequência com que estes tipos ocorrem entre Os Sertões e A Margem da História é aproximadamente a mesma, gráfico da Figura 37 exhibe um resumo desta afirmação.

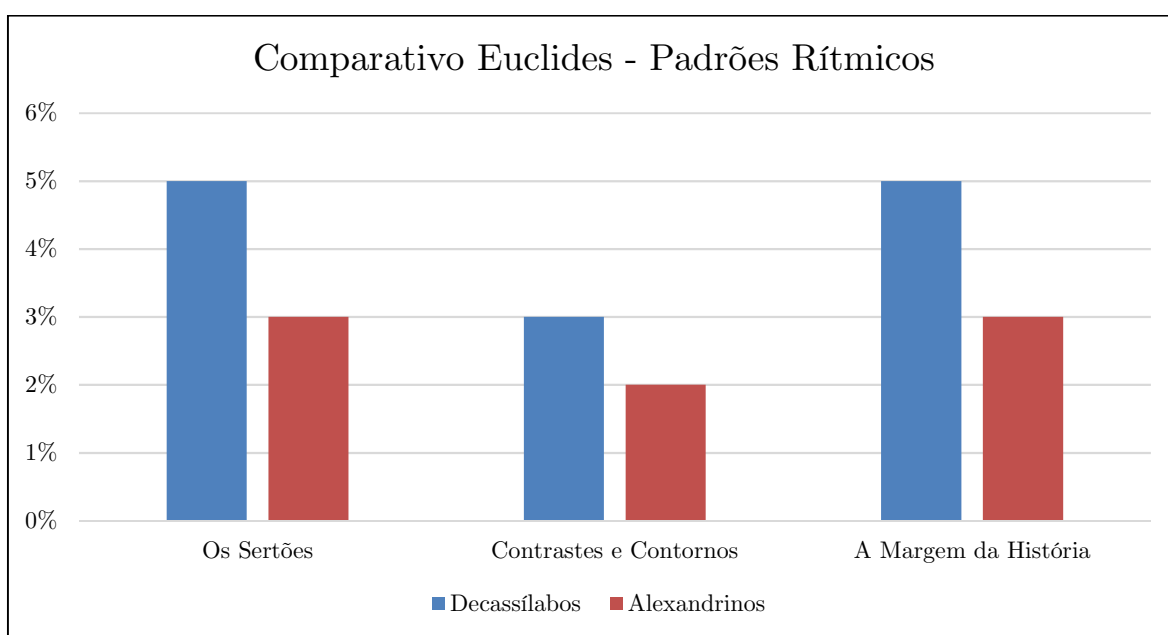


Figura 37: Comparativo dos Padrões Rítmicos sobre Prosa de Euclides da Cunha

Outro fator analisado, foram as distâncias entre os segmentos com estrutura métrica. Esses valores de distâncias são medidos calculando quantos segmentos fráscicos (sem métrica) separam as estruturas de versificação. Na Tabela 26, em “distância entre sentenças” também são apresentadas informações sobre os valores médios de distância e desvio padrão ao longo das obras analisadas para sentenças completas, início e fim de sentenças com estrutura métrica.

O desvio padrão evidencia o comportamento relacionado à distribuição das estruturas metrificadas ao longo do texto. Os altos valores relacionados ao desvio padrão evidenciam uma dispersão irregular das estruturas identificadas nos textos. Ou seja, não há, nas obras analisadas, um padrão regular de distribuição de estruturas, em intervalos bem definidos.

Os experimentos realizados até este ponto mostraram a eficiência do MIVES quanto ao processo e identificação das sentenças com estrutura métrica em texto em prosa. O comparativo entre os resultados relacionados a métrica com trabalho realizado manualmente por Campos[1996] mostraram o quanto a eficiência da metodologia adotada. Além disso, os resultados apresentados neste documento dos experimentos realizados com as obras *Contrates e Contornos* e em *A Margem da História* demonstram que padrões relacionados a sentenças com estrutura métrica se repetiram também nestas obras. Não faz parte dos nossos objetivos afirmar que o comportamento versificatório de Euclides se repete nestas obras, mas os padrões identificados pelo MIVES nos permitem supor que realmente se repetem.

Tabela 26: Dados Comparativos entre as Obras de Euclides

Parâmetros	Obra		
	Os Sertões	Á Margem da História	Contrastes e Contornos
Total de Sentenças	8564	1066	1598
Sentenças completas com estrutura métrica (total e percentual)	652 7,61%	76 7,13%	82 5,13%
Início de sentenças com estrutura métrica (total e percentual)	1746 20,39%	227 21,29%	282 17,65%

Finais de sentenças com estrutura métrica (total e percentual)	1728 20,18%	219 20,54%	235 14,71%
Quantidade de Decassílabos (sentenças, trechos iniciais e finais)	354 859 739 Total: 1952	36 93 81 Total: 210	43 134 92 Total: 269
Quantidade de Dodecassílabos (sentenças, trechos iniciais e finais)	360 854 1081 Total: 2295	34 106 116 Total: 256	40 156 145 Total: 341
Distância entre sentenças (média, desvio padrão, mínimo e máximo)	12,63±12,88 (0-82) 4,02±4,54 (0-46) 3,76±4,13 (0-30)	14,62±16,89(0-77) 4,33±4,37(0-27) 4,49±4,64(0-23)	18,38±19,61(0-109) 4,71±5,60(0-35) 5,85±6,61(0-59)

4.4 Análise de Outras Obras da Prosa Literária Brasileira.

Com o objetivo de expandir nossas análises e testar as capacidades do MIVES de realizar buscas por estruturas de versificação em textos em prosa, realizamos experimentos utilizando outras obras da literatura brasileira. Além disso, pretende-se descobrir se os padrões identificados nas obras analisadas de Euclides se repetem dentre obras de outros autores do mesmo período. Selecionamos três obras de diferentes autores: Canaã de Graça Aranha, Recordações do Escrivão Isaías Caminha de Lima Barreto, Dom Casmurro de Machado de Assis. Todos os experimentos realizados nesta seção seguiram o mesmo ambiente dos experimentos anteriores. Estas obras foram escolhidas por terem sido publicadas durante o mesmo período da literatura brasileira.

4.4.1 Graça Aranha - Canaã

Publicada pela primeira vez no Brasil em 1902 e escrita por João Pereira de Graça Aranha, a obra está posicionada entre as grandes obras da literatura brasileira. O romance aborda a imigração Alemã para o Espírito Santo por meio da relação entre dois personagens. Como já descrito nesta dissertação, não encontramos qualquer referência a estudos de mesma natureza, a aquelas realizadas por Augusto de Campos sobre a prosa de Os Sertões, para outras obras. Por não sabermos previamente quais metros iríamos encontrar e para manter a uniformidade entre os experimentos, configuramos o MIVES para buscar por possíveis estruturas de versificação com 10 a 12 sílabas poéticas. Em seguida, analisamos os resultados para verificar quais dos metros apareceriam com maior frequência e como estes estavam distribuídos ao longo do texto.

A primeira etapa foi dedicada a busca de estruturas de versificação formadas por sentenças completas. Após processar 4273 sentenças o MIVES retornou 388 estruturas de versificação com métrica variada. Nessa etapa do experimento não encontramos uma população significativa dentre os metros buscados, um tipo específico de metro sobre os segmentos escandidos, a ponto de traçar um perfil preferencial sobre a prosa de Canaã. Foram 164 decassílabos, 175 hendecassílabos e 179 dodecassílabos. O gráfico da Figura 38 mostra a distribuição da frequência absoluta das estruturas com estrutura métrica sobre um intervalo de 200 sentenças. O da Figura 39 como estão distribuídas estas sentenças ao longo do texto. Analisando as informações dos dois gráficos é possível verificar que existem zonas maior densidade e alguns intervalos onde a frequência diminui (Figura 38) acompanhada por uma dispersão maior entre as sentenças.

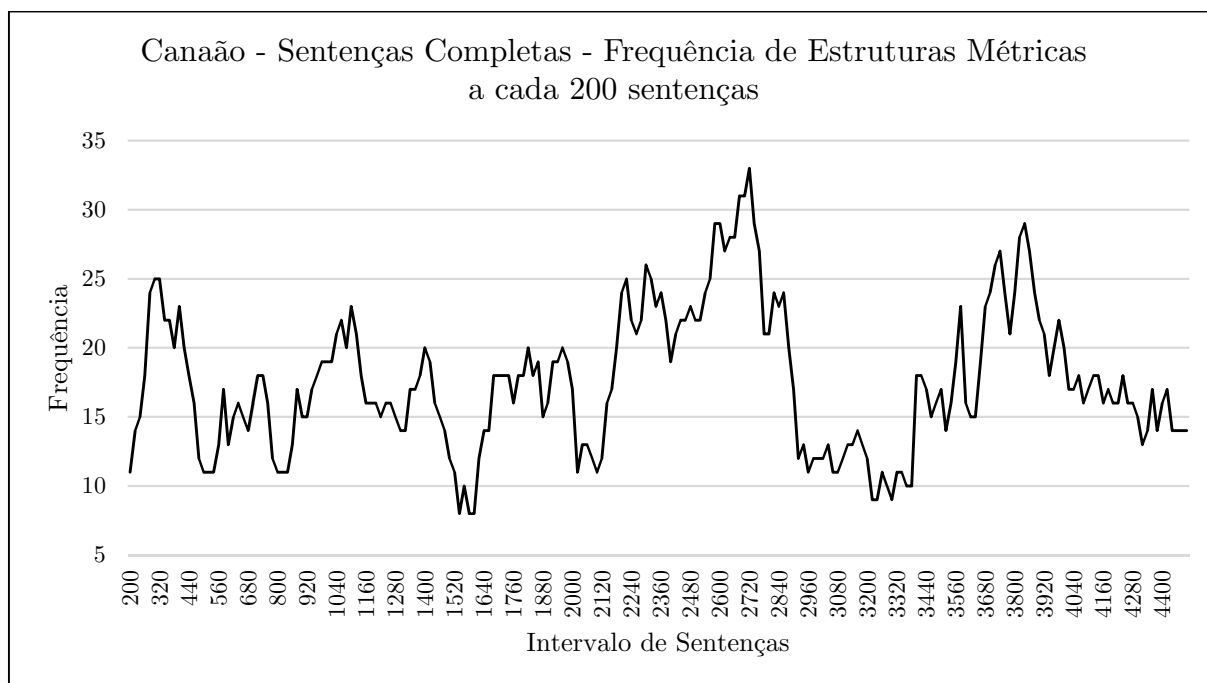


Figura 38: Canaã – Sentenças Completas - Distribuição da Frequência Absoluta.

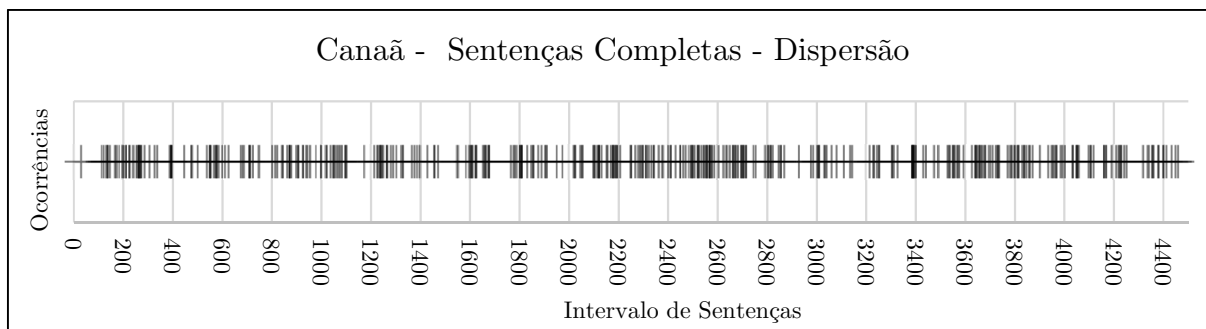


Figura 39: Canaã – Sentenças Completas - Distribuição das sentenças Métricas ao longo do texto

Existem sobre as estruturas até este ponto identificadas padrões métricos. Para justificar esta afirmação, além da quantidade de deca e dodecassílabos apresentados, e com o objetivo de ampliar esta análise, identificamos os principais Posicionamento de Tônicas sobre as estruturas identificadas. O resultado pode ser observado no gráfico da Figura 40, dentre os padrões rítmicos conhecidos da poesia estão 11 heroicos, 4 martelos, 3 gaitas galega, 2 sáficos e cerca de 18 alexandrinos. Nesta etapa dos experimentos com Canaã o MIVES reportou 359 variações rítmicas, de forma geral uma lista com os 10 Posicionamento de Tônicas com maior frequência está disponível na Tabela 27. A Figura 18 apresenta alguns exemplos de estruturas de versificação sobre sentenças completas encontradas em Canaã, com suas respectivas formas de escansão apresentadas pelo MIVES.

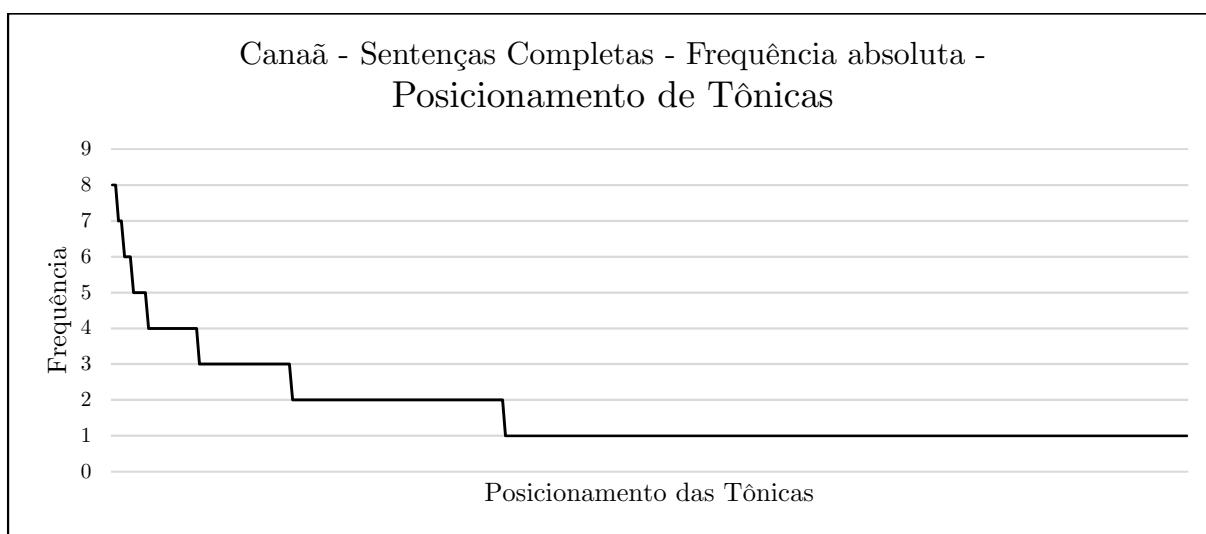


Figura 40: Canaã - Sentenças Completas- Frequência Absoluta - Posicionamento das Tônicas.

Tabela 27: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentença Completas

Posicionamento de Tônicas	Frequência
2 4 7 10	8
2 4 7 11	8
2 6 10	7
3 7 11	7
2 5 10	6
1 4 6 10	6

2 4 6 8 10	6
2 6 9 12	5
3 7 10	5
2 6 8 10	5

Tabela 28: Exemplo de Sentenças Completas – Canaã

Sentença Completa com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Toda ela era a própria indolência.	To/da/e/la/e/ra a/ pró /pria in/do/ lên /cia.	10
	To/da/e/la/e/ra/ a/ pró /pria in/do/ lên /cia.	11
	To/da/e/la/e/ra/ a/ pró /pria/ in/do/ lên /cia	12
E os seus olhos tristes obscureceram-se.	E os/ seus/ o /lhos/ tris /tes/ obs/cu/re/ ce /ram-/se.	10
	E/ os/ seus/ o /lhos/ tris /tes/ obs/cu/re/ ce /ram-/se.	11
Sentaram-se os dois junto à janela aberta.	Sen/ ta /ram/se/ os/ dois / jun /to à/ ja/ ne /la a/ ber /ta.	11
	Sen/ ta /ram/se/ os/ dois / jun /to à/ ja/ ne /la a/ ber /ta.	12
Sempre este amarelo a nos perseguir...	Sem /pre es/te a/ma/ re /lo a/ nos/ per/se/ guir ...	10
	Sem /pre/es/te a/ma/ re /lo a/ nos/ per/se/ guir ...	11
	Sem /pre/es/te/ a/ma/ re /lo a/ nos/ per/se/ guir ...	12
Questão de amor, ou antes questão de consciência...	Ques/ tão / de a/ mor ,/ ou an /tes/ ques/ tão / de/ cons/ci/ ên /cia...	12

Seguindo com o experimento, buscamos por estruturas de versificação constituídas por início de sentenças, empregando a mesma faixa de metros utilizada no experimento anterior, de 10 a 12 sílabas poéticas. Nesta etapa MIVES retornou 951 estruturas de versificação. Destes 396 decassílabos, 548 hendecassílabos e 478 dodecassílabos. O gráfico da Figura 41 mostra quais as concentrações mais densas de sentenças em intervalos de 200 segmentos enquanto o gráfico da Figura 42 mostra a dispersão destas estruturas ao longo da obra. No primeiro gráfico é possível visualizar que existem regiões com altas concentrações, a maior com 56 enquanto a menor com 26 ocorrências de sentenças com estrutura métrica, embora diferença seja quase 50%, no restante da obra esse comportamento de altas amplitudes não se repete. Em um total de 216 intervalos computados, a média das frequências é 42 com um desvio padrão destes valores de $\pm 5,0$ mostrando que as amplitudes nos valores de frequência nos intervalos não são muito distantes.

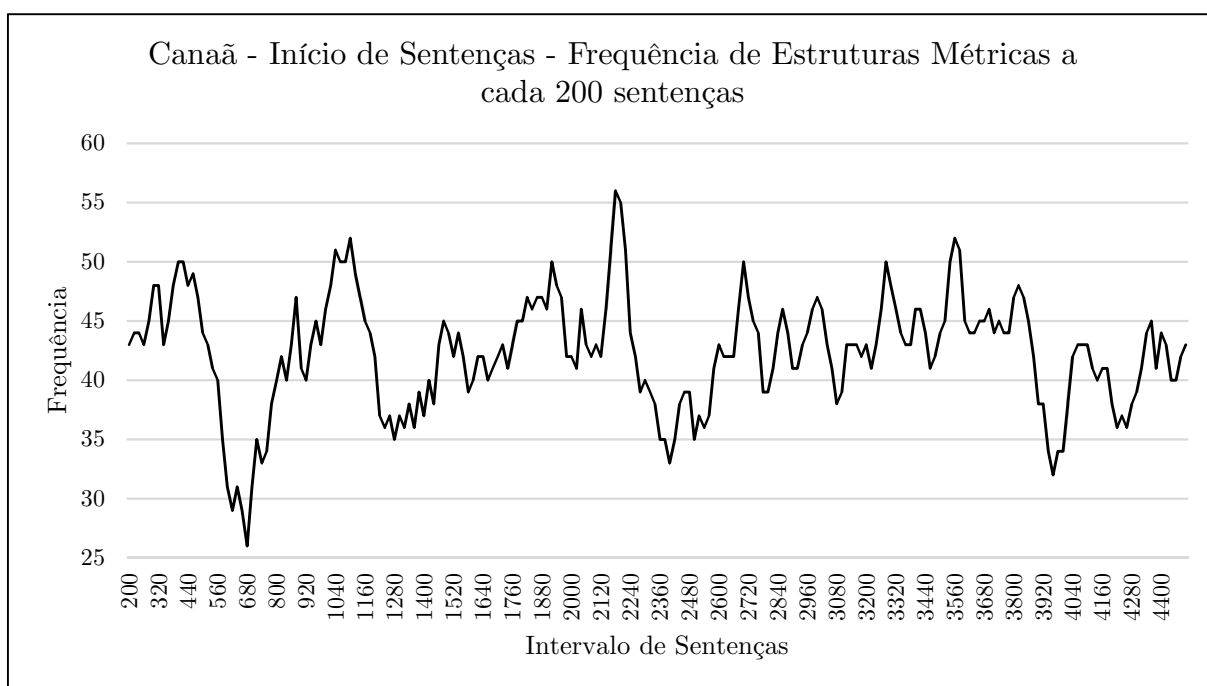


Figura 41: Canaã – Início de Sentenças - Distribuição de Frequências Absolutas.

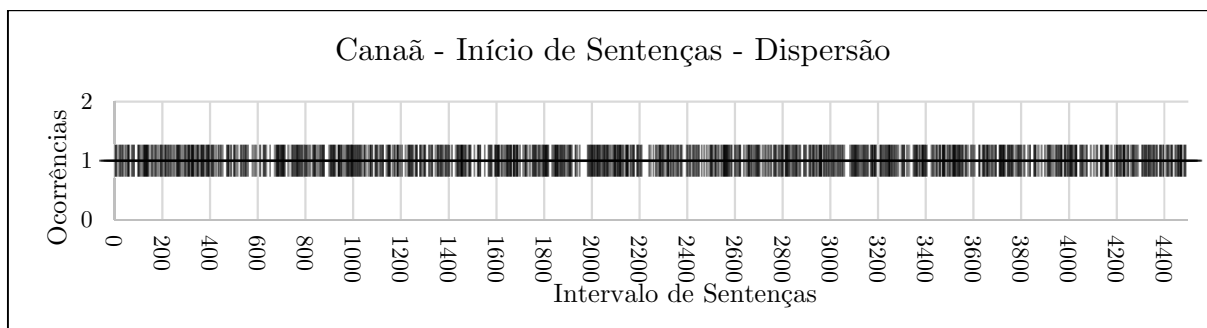


Figura 42: Canaã – Início de Sentenças - Distribuição das sentenças métricas ao longo do texto

Em relação da dispersão entre as sentenças, informação apresentada pelo gráfico da Figura 42, aparentemente existem regiões de grandes concentrações com poucos

intervalos ente as sentenças com estrutura métrica, o maior valor de distância entre sentenças encontrado foi de 33 sentenças, mas média de distância é de apenas 4 mostrando que de fato existe uma proximidade entre os inícios de sentenças com estrutura métrica identificadas pelo MIVES em Canaã. Dentre um universo de 558 variações de posicionamento de tônicas destacamos a presença de 26 heroicos, 12 Martelos, 6 Gaitas Galega, 4 Sáficos e 39 alexandrinos, padrões rítmicos característicos da poesia. De forma geral, o gráfico da Figura 43 apresenta o comportamento relacionado a frequência absoluta com que os padrões de posicionamento de tônicas aparecem no texto, enquanto a Tabela 29 apresenta uma relação dos posicionamentos de tônicas com maior frequência nos inícios de sentença com estrutura métrica de Canaã.

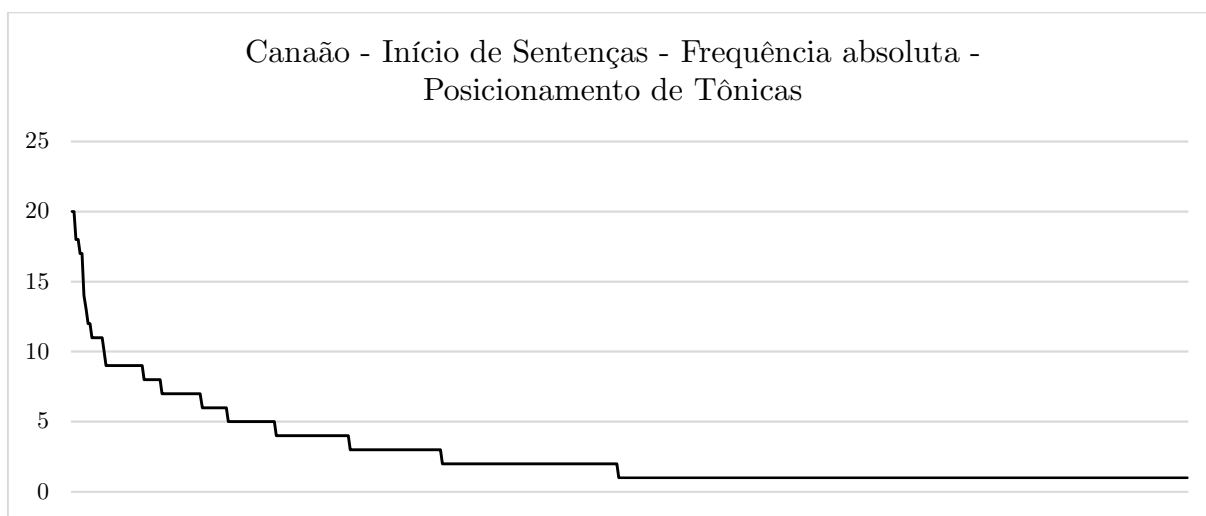


Figura 43: Canaã - Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Tabela 29: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentença

Posicionamento de Tônicas	Frequência
2 5 10	20
2 6 10	20
2 5 8 11	18
2 4 7 11	18
3 7 11	17
2 4 7 10	17
3 7 10	14
2 5 8 12	13
2 5 8 10	12
3 6 10	12

A Tabela 30 apresenta uma relação com alguns exemplos de versos extraídos de Canaã neste experimento.

Tabela 30: Exemplos de Estruturas de Início de Sentença - Canaã

Início de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Tudo era um abandono preguiçoso,	Tu /do e /ra um/ a/ban/ do /no/ pre/gui/ ço /so,	10
	Tu /do/ e /ra um/ a/ban/ do /no/ pre/gui/ ço /so,	11
	Tu /do/ e /ra/ um/ a/ban/ do /no/ pre/gui/ ço /so,	12
E, sorrindo malicioso, ergueu-se,	E,/ so/ rrin /do/ ma/li/ci/ o /so, er/ gheu -/se,	10
	E,/ so/ rrin /do/ ma/li/ci/ o /so,/ er/ gheu -/se,	11
Para o norte, para o sul, para o poente,	Pa /ra o/ nor /te,/ pa /ra o/ sul ,/ pa /ra o/ po/ en /te,	11
	Pa /ra/ o/ nor /te,/ pa /ra o/ sul ,/ pa /ra o/ po/ en /te,	12
Numa devoção alegre e radiante,	Nu /ma/ de/vo/ ção / a/ le /gre e/ ra/di/ an /te,	11
	Nu /ma/ de/vo/ ção / a/ le /gre/ e/ ra/di/ an /te,	12
A ordem não é um princípio moral;	A or /dem/ não / é/ um/ prin/ cí /pio/ mo/ ral ;	10
	A/ or /dem/ não / é/ um/ prin/ cí /pio/ mo/ ral ;	11
A liberdade é como a própria vida,	A/ li/ber/ da /de/ é/ co /mo a/ pró /pria/ vi /da,	11
	A/ li/ber/ da /de/ é/ co /mo/ a/ pró /pria/ vi /da,	12

Na última etapa buscamos por estruturas de versificação constituídas sobre final de sentenças. Desta vez foram retornadas 956 estruturas de versificação, formados por 362 decassílabos, 570 hendecassílabos e 574 dodecassílabos. A Tabela 31 apresenta alguns exemplos das estruturas encontradas nesta experiência.

Tabela 31: Exemplos de Final de Sentenças - Canaã

Final de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
a uma plena expansão da individualidade.	au/ma/ ple /na ex/pan/ são / da in/di/vi/dua/li/ da /de.	12
que prolongavam a ilusão da madrugada.	que/ pro/lon/ ga /vam/ a i/lu/ são / da/ ma/dru/ ga /da.	12
enquanto ela rangia com um grito agudo.	en/ quan /to e/la/ ran/ gia / com/ um/ gri /to a/ gu /do.	11
	en/ quan /to e /la/ ran/ gi /a/ com/ um/ gri /to a/ gu /do.	12
bufando e cantando insofridos a erva.	bu/ fan /do e/ can/ tan /do in/so/ fri /dos/aer /va.	10
	bu/ fan /do/ e/ can/ tan /do in/so/ fri /dos/aer /va.	11
	bu/ fan /do/ e/ can/ tan /do/ in/so/ fri /dos/aer /va.	12
piando como pássaros da morte.	pi/ an /do/ co /mo/ pá /ssa/ros/ da/ mor /te.	10
erguendo indolente o sombreiro de palha.	er/ guen /do in/do/ len /te o/ som/ brei /ro/ de/ pa /lha.	11
	er/ guen /do in/do/ len /te o/ som/ bre /i/ro/ de/ pa /lha.	12

Analisando o gráfico apresentado pela Figura 44, é possível concluir que não existem uma região com alta concentração de sentenças com estrutura métrica, embora o gráfico mostre picos aparentes que podem ser interpretados como uma alta concentração de sentenças com estrutura métrica, o gráfico da Figura 45 mostra que estas estruturas, embora não seja uniforme, estão bem distribuídas ao longo do texto. A distância média entre as sentenças com estrutura métrica é de aproximadamente 4, enquanto a maior distância é de 35 com um desvio padrão de $\pm 4,2$.

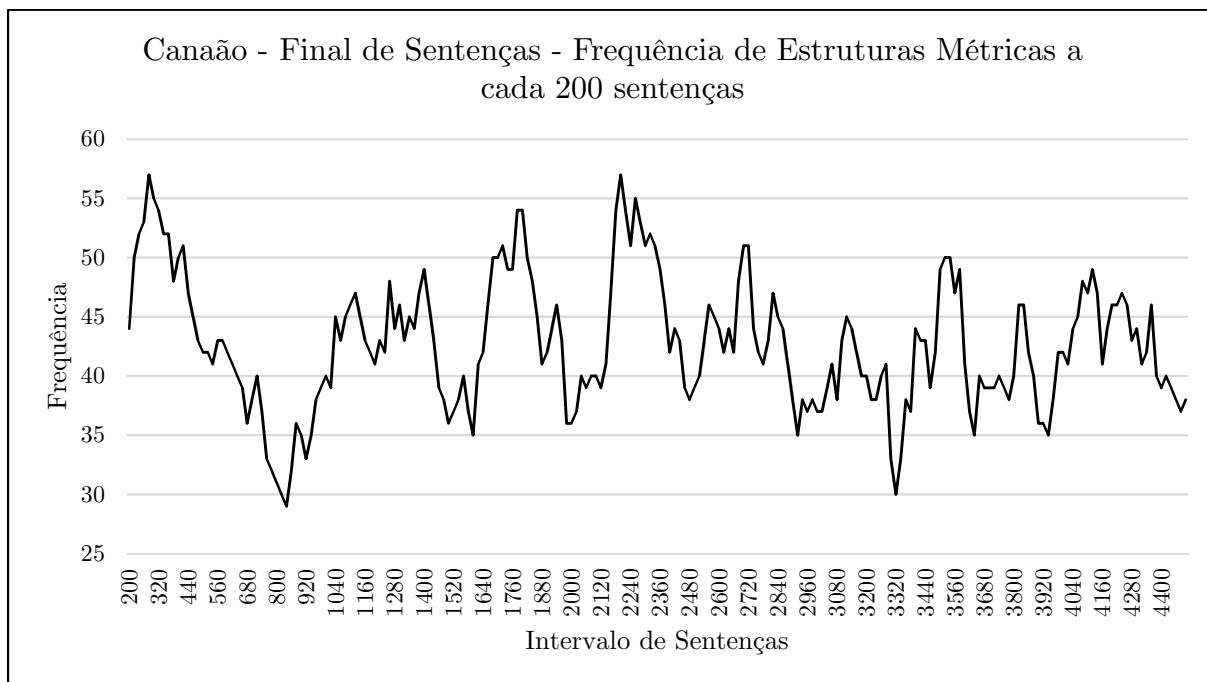


Figura 44: Canaã – Final de Sentença - Distribuição da Frequências Absoluta.

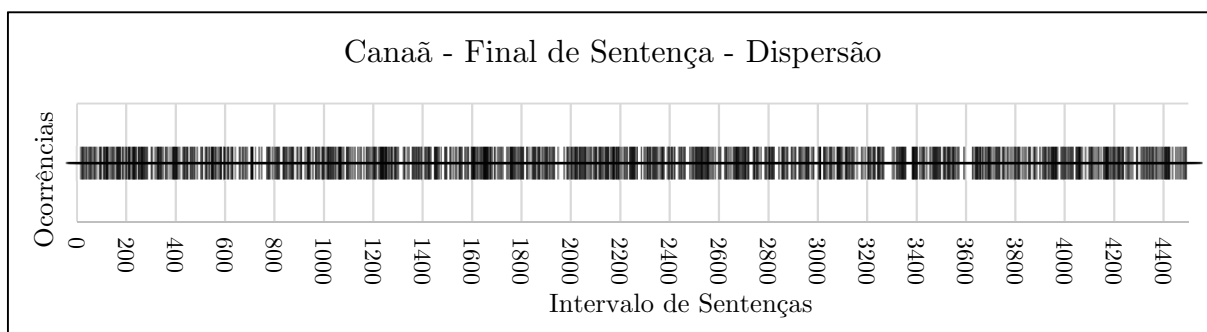


Figura 45: Canaã – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto.

Além de analisarmos como estão distribuídas as sentenças com estrutura métrica ao longo do documento verificamos quais as seriam as combinações ou padrões de tônicas presentes no texto. O resultado pode ser observado no gráfico da Figura 46 onde podemos concluir que existem padrões que ocorrem com maior frequência, 25 vezes, enquanto outras tem o número de ocorrência igual a 1. A partir de um total de 602 variações, a Tabela 32 apresenta um ranking das 10 combinações de tônicas que aparecem com maior frequência. Dentro desta população total, os padrões rítmicos clássicos em Canaã destacam-se os 17 heroicos, 10 martelos, 7 gaitas galega, 3 sáficos e 48 alexandrinos.

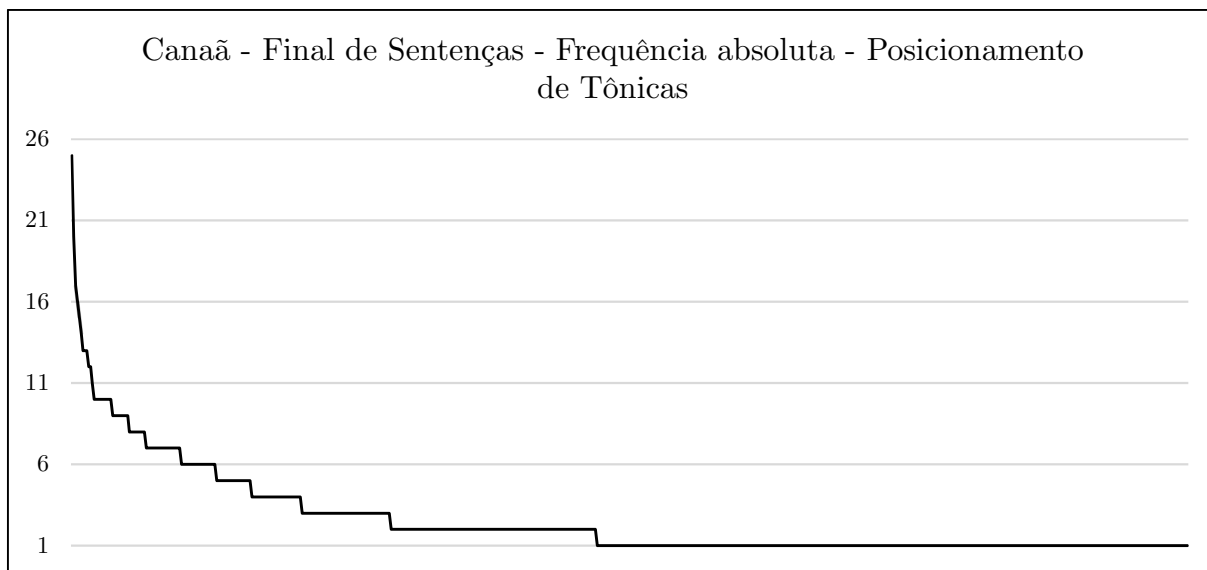


Figura 46: Canaã - Final de Sentenças- Frequência Absoluta - Posicionamento das Tônicas

Tabela 32: Canaã - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentença

Posicionamento de Tônicas	Frequência
2 5 8 11	25
3 7 11	20
2 4 7 10	17
3 7 10	16
4 7 11	15
3 6 8 11	14
2 4 8 11	13
1 3 7 11	13
2 4 7 11	13
2 5 7 10	12

Nesse experimento identificamos inúmeras estruturas de versificação com destaque para aquelas que possuem métrica entre 10 e 12 sílabas por representarem expressivamente, a maioria dos metros encontrados. A Tabela 33 apresenta um resumo dos resultados obtidos neste experimento.

Tabela 33: Resumo das Experiências com Canaã

Parâmetros	Resultado
Total de Sentenças	4273
Sentenças completas com estrutura métrica (total e percentual)	388 9,08%
Início de sentenças com estrutura métrica (total e percentual)	951 22,26%

Finais de sentenças com estrutura métrica (total e percentual)	956 22,37%
Quantidade de Decassílabos (sentenças, trechos iniciais e finais)	188 469 362 Total: 1019
Quantidade de Dodecassílabos (sentenças, trechos iniciais e finais)	201 453 574 Total: 1228
Distância entre sentenças (média, desvio padrão, mínimo e máximo)	10,49±13,60 (0-88) 3,72±4,04 (0-33) 3,70±4,24 (0-35)

Resumindo as três etapas deste experimento, podemos concluir que na obra de Graça Aranha existem diversas estruturas de versificação de diferentes metros. Nas sentenças completas não foi identificado pelo MIVES uma população de metros que se destacasse em número significativo. Porém, na etapa em que buscamos por estruturas de versificação de início e fim de sentenças, os metros em maior número, entre 10 e 12, tiveram um destaque por se apresentarem em maior número, dentre os metros encontrados. Foram encontrados, nas três etapas do experimento, padrões rítmicos característicos da poesia brasileira, dentre heroicos, martelos, sáficos, gaitas galega e alexandrinos, foram cerca de 203 ocorrências. Outro fator importante, apresentado pela Tabela 33, são os valores de média e desvio padrão relacionados a distância entre as sentenças com estrutura métrica. O valor apresentado para estes parâmetros nos permite concluir que não existe um padrão de comportamento definido para quando essas estruturas podem ocorrer na obra analisada neste experimento.

4.4.2 Lima Barreto - Recordações do Escrivão Isaías Caminha

Escrito por Afonso Henriques de Lima Barreto e publicado em 1917, *Recordações do Escrivão Isaías Caminha*, tem como tema racismo e subordinação, é um romance em que seu enredo gira em torno de um jovem negro, nascido no interior do Rio de Janeiro, que se vê insatisfeito frente às condições de sobrevivências e críticas voltadas para pessoas da sua raça, e então vai luta, saindo do interior e indo morar na capital em busca de realização pessoal.

Submetemos esta obra de Lima Barreto ao MIVES seguindo os mesmos parâmetros e estratégias a que submetemos *Canaã*, utilizamos os mesmos parâmetros de escansão e buscamos pela mesma faixa de metros, entre 10 e 12 sílabas poéticas. Na primeira etapa, na busca por estruturas de versificação constituídas por sentenças completas, o sistema retornou cerca de 301 estruturas de versificação. Sendo 149 decassílabos, 142 hendecassílabos e 146 dodecassílabos. A Tabela 34 apresenta alguns exemplos de estruturas de versificação identificadas pelo MIVES neste experimento.

Tabela 34: Exemplo de Sentenças Completas - Recordações do Escrivão Isaias Caminha

Sentença Completas com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Pelo caminho, viemos, os dois, calados.	Pe/lo/ ca/mi/nho,/ vi/e/mos,/ os/dois,/ ca/la/dos.	12
O trem parara e eu abstinha-me de saltar.	O/trem/ pa/ra/ra e eu /abs/ti/nha/me/de/ sal/tar.	12
Demais, a emanação da minha pessoa.	De/mais,/ a e/ma/na/ção/ da/mi/nha/pe/sso/a.	11
	De/mais,/ a/ e/ma/na/ção/ da/mi/nha/pe/sso/a.	12
O próprio Oliveira, tão parvo e tão besta.	O/pró/prio O/li/vei/ra,/tão/par/vo e/tão/bes/ta.	11
	O/pró/prio O/li/ve/i/ra,/tão/par/vo e/tão/bes/ta.	12
Havia barrancos, covas, capinzais...	Ha/via/ ba/rran/cos,/co/vas,/ ca/pin/zais...	10
	Ha/vi/a/ ba/rran/cos,/co/vas,/ ca/pin/zais...	11

Seguindo mesma estratégia adotada nos experimentos anteriores analisamos o comportamento relacionado a possíveis concentrações de sentenças com estrutura métrica e de que forma estão distribuídas ao longo do texto. O resultado pode ser observado no gráfico da Figura 47. Neste gráfico fica clara a presença de regiões com alta concentração de sentenças métricas em relação a outras. O que pode ser comprovado pelas distâncias apresentadas entre as sentenças com estrutura métrica apresentadas pelo gráfico da Figura 48. Nesta etapa, identificamos que a maior distância entre as sentenças é de até 95 sentenças sem a presença de métrica. Também neste experimento analisamos quais padrões de posicionamento de tônicas ou padrões rítmicos se repetiam com mais frequência. Uma lista com os 10 padrões ritmos que mais se repetiram nas sentenças completas com estrutura métrica da obra de Lima Barreto consta na Tabela 35 enquanto o comportamento geral das frequências é apresentado pelo gráfico da Figura 49. Dentre 320 variações de posicionamento de tônicas, os padrões rítmicos mais comuns na poesia encontrados pelo MIVES se resumem em 7

heroicos, 5 martelos, 5 gaitas galega e 6 alexandrinos.

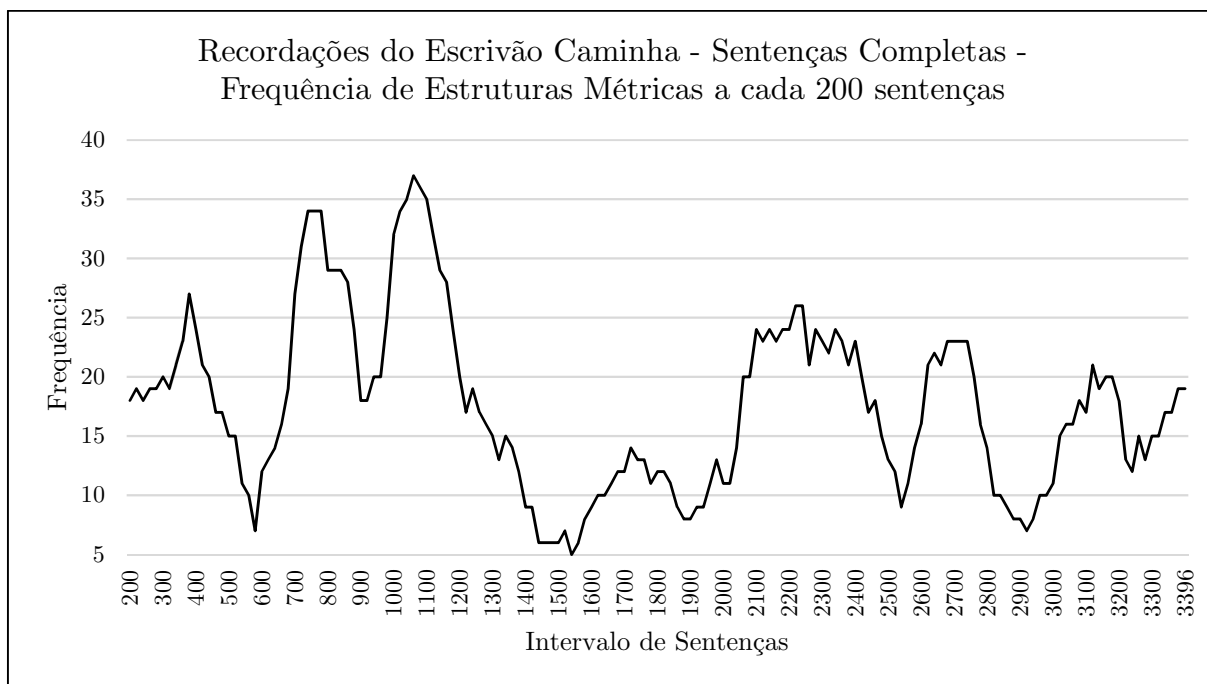


Figura 47: Recordações do Escrivão Caminha – Sentenças Completas - Distribuição da Frequência Absoluta.

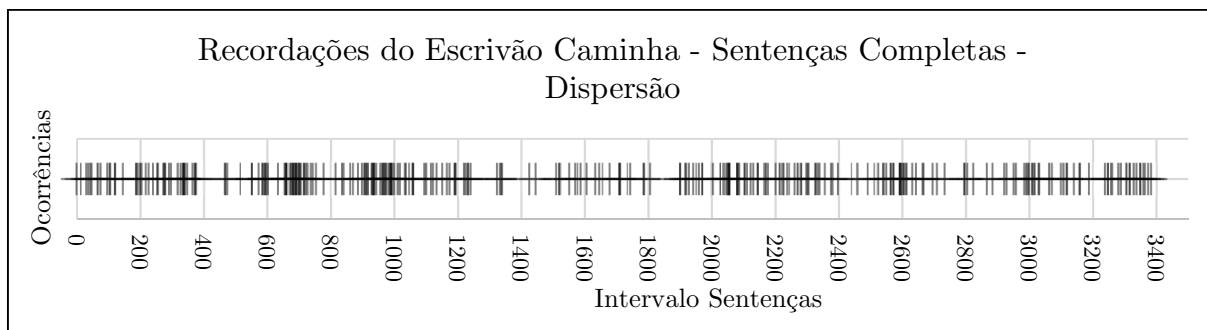


Figura 48: Recordações do Escrivão Caminha – Sentenças Completas - Distribuição das sentenças métricas ao longo do texto

Tabela 35: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentenças Completas

Posicionamento de Tônicass	Frequência
2 4 7 11	6
2 6 10	5
1 4 7 10	5
3 6 10	5
2 7 10	5
4 7 10	5
2 4 7 10	5
1 4 6 8 10	4

1 5 7 10 12	4
2 5 7 11	4

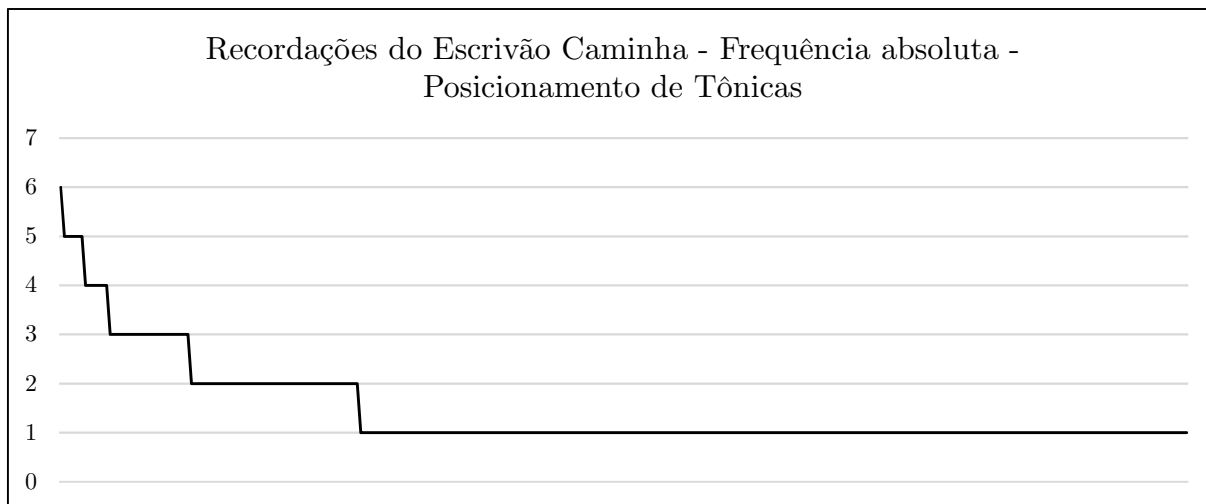


Figura 49: Recordações do Escrivão Caminha – Sentenças Completas - Frequência Absoluta - Posicionamento das Tônicas

Passamos para segunda etapa deste experimento, buscamos por início de sentenças com estrutura métrica. Nesta etapa o MIVES identificou 768 estruturas de versificação. Sendo 400 decassílabos, 446 hendecassílabos e 344 dodecassílabos. A Tabela 36 apresenta alguns exemplos de estruturas de versificação encontradas pelo MIVES nesta etapa do experimento.

Tabela 36: Exemplo de Início de Sentenças - Recordações do Escrivão Isaías Caminha

Início de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Encostado ao espaldar do banco,	En/cos/ ta /do/ ao/ es/pal/ dar / do/ ban /co,	10
E a facilidade com que os aceitava,	E a/ fa/ci/li/ da /de/ com/ que os/ a/ cei /ta/va,	10
Como se fosse um furto, um crime,	Co /mo/ se/ fo /sse/ um/ fur /to,/ um/ cri /me,	10
Aferrara-me a duas amarras,	A/fe/ r ra/ra-/me/ a/ du /as/ a/ ma /rras,	10
A chuva começou a cair fracamente,	A/ chu /va/ co/me/ ç ou a/ ca/ ir / fra/ca/ men /te,	11

	A/ chu /va/ co/me/ ço /u a/ ca/ ir / fra/ca/ men /te,	11
Entretanto, quantas dores, quantas angústias!	En/tre/ tan /to,/ quan /tas/ do /res,/ quan /t as/ an/ gús /tias!	12

Nesta etapa, em que estamos buscamos por início de sentenças com estrutura métrica, analisamos também como estas estruturas estão distribuídas ao longo do texto. Para este caso, o gráfico da Figura 50, nos permite concluir que existe regiões com grandes concentrações com uma diferença entre um máximo (69) e mínimo (33) de 36 sentenças, apesar de apresentar visualmente grandes amplitudes em algumas regiões, o gráfico de dispersão da Figura 51 nos mostra que estas sentenças, embora não seja uniforme, estão distribuídas ao longo de todo o texto.

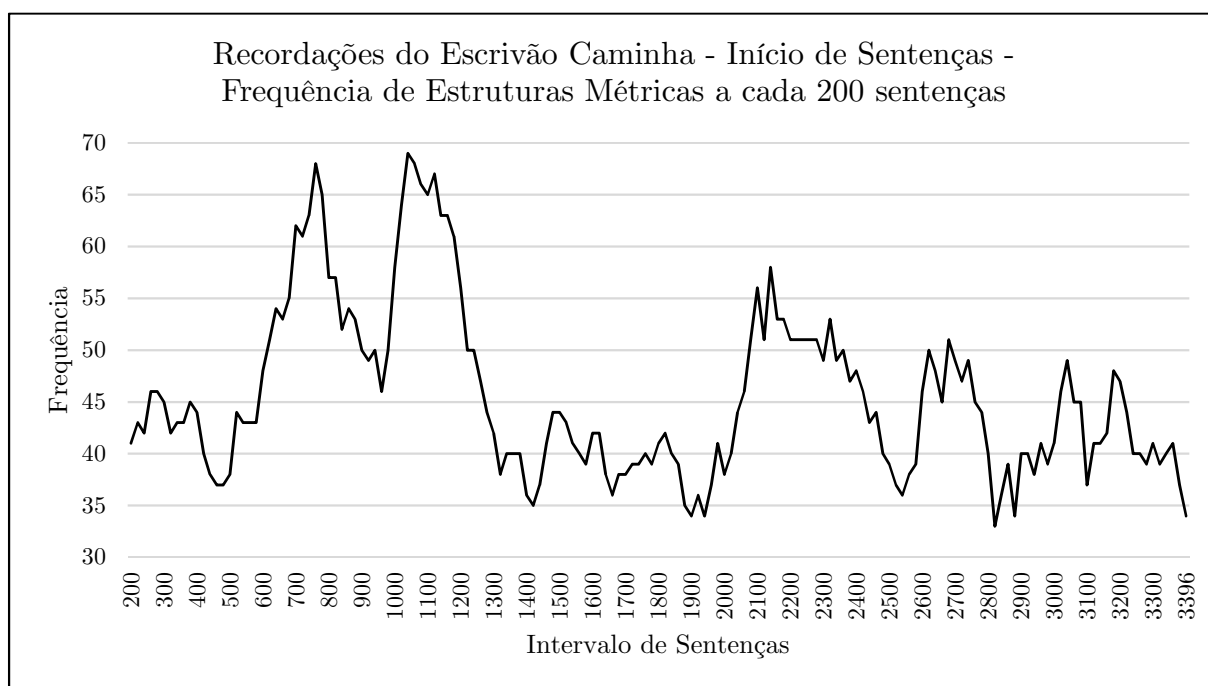


Figura 50: Recordações do Escrivão Caminha – Início de Sentenças - Distribuição de Frequências Absolutas.

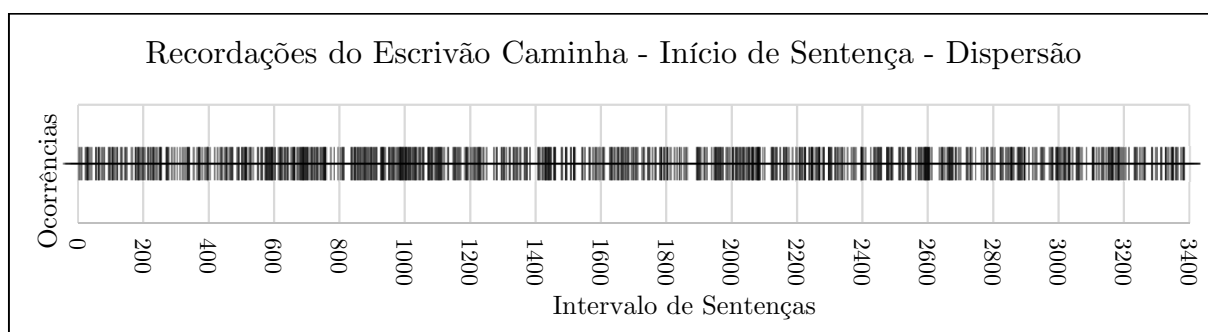


Figura 51: Recordações do Escrivão Caminha – Início de Sentenças - Distribuição das sentenças

Métricas ao Longo do Texto.

Além disso, analisamos a frequência dos padrões rítmicos nesta etapa do experimento. Uma listagem das dez variações de posicionamento de tônicas mais comuns, de um total de 527 variações, pode ser vista na Tabela 37, enquanto o comportamento geral destas frequências no gráfico da Figura 52. Dentre os padrões rítmicos característicos da poesia destacamos a ocorrência de 17 heroicos, 16 martelos, 9 gaitas galega, 1 Sáfico e 17 alexandrinos.

Tabela 37: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentenças

Posicionamento das Tônicas	Frequência
1 4 7 10	16
3 6 10	16
2 5 8 12	15
2 5 7 11	14
2 5 8 11	14
2 5 7 12	13
2 6 10	13
2 4 7 11	13
2 5 7 10	12
2 5 10	12

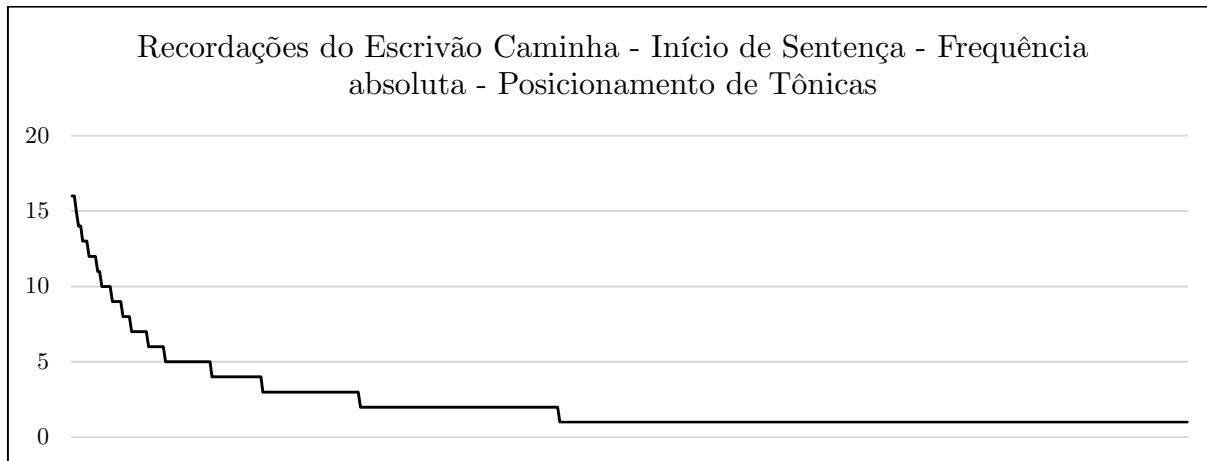


Figura 52: Recordações do Escrivão Caminha – Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Continuando nosso experimento realizamos a busca por finais de sentenças com estrutura métrica. Ao final do processamento, o MIVES retornou 755 estruturas de versificação. Foram 314 decassílabos, 437 hendecassílabos e 426 dodecassílabos. A Tabela 38 apresenta algumas das estruturas de versificação apresentadas como resultado pelo MIVES.

Tabela 38: Exemplo de Estruturas de Final de Sentenças - Recordações do Escrivão Isaías Caminha

Final de Sentença	Escansão segundo MIVES	Nº
-------------------	------------------------	----

com Estrutura Métrica		Sílabas Poéticas
deram-me anseios de inteligência.	de/ram/me an/se/i/os/ de in/te/li/gên/cia.	10
fantásticas, discordantes e fugidias..."	fan/tás/ti/cas,/ dis/cor/dan/tes/ e/ fu/gi/di/as...	12
cujo aroma rescendia e tentava a fumá-los.	cu/jo a/ro/ma/ res/cen/dia/ e/ ten/ta/va a/fu/má/los.	12
na bruma do futuro fugidio.	na/bru/ma/ do/ fu/tu/ro/ fu/gi/di/o.	10
um verme asqueroso e um réptil nojento.	um/ver/me as/que/ro/so e um/rép/til/no/jen/to.	10
	um/ver/me/ as/que/ro/so e um/rép/til/no/jen/to.	11
	um/ver/me/ as/que/ro/so/ e um/rép/til/no/jen/to.	12
esquinas, as vendas estavam cheias.	es/qui/nas,/ as/ven/das/ es/ta/vam/chei/as.	10
	es/qui/nas,/ as/ven/das/ es/ta/vam/che/i/as.	11

Seguindo com nossas análises é possível identificar através do gráfico apresentado pela Figura 53, possuem áreas com uma maior concentração de estruturas métricas com intervalos onde estas ocorrem com menor frequência. A diferença a maior concentração (63) e a menor (23) é de 23.

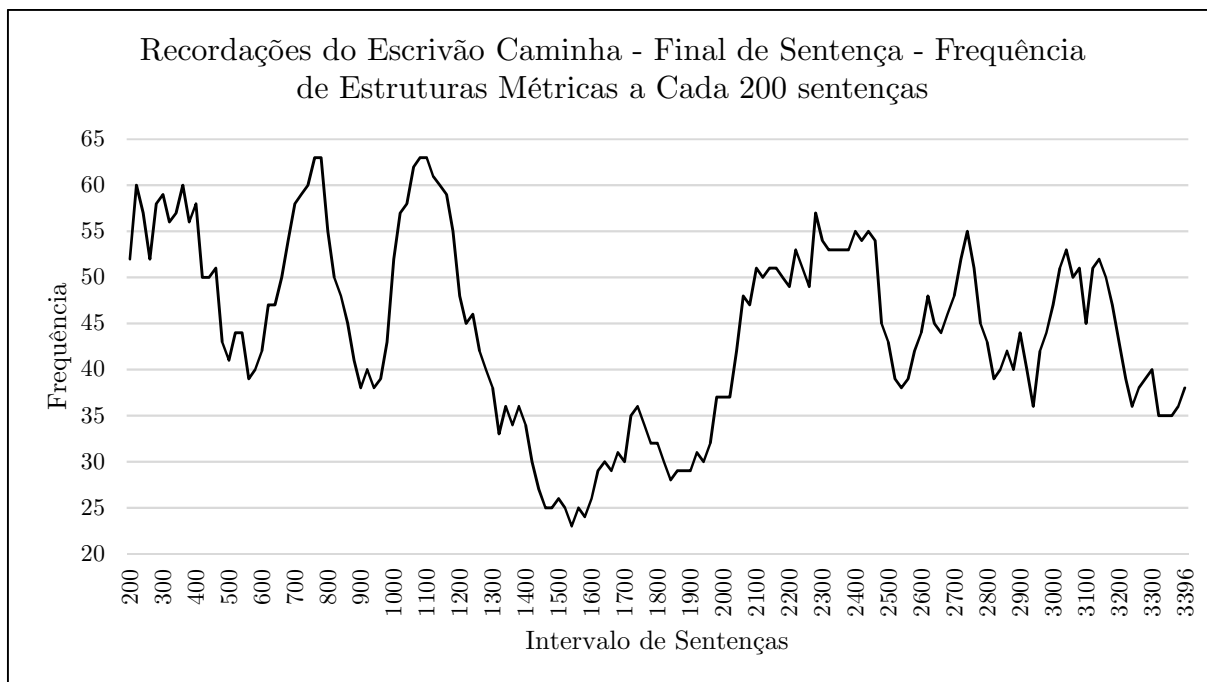


Figura 53: Recordações do Escrivão Caminha – Final de Sentença - Distribuição de Frequência Absoluta.

A dispersão ou distribuição destas sentenças ao longo do texto pode ser observado pelo gráfico da Figura 54. Através do gráfico é possível visualizar que as sentenças com estrutura métrica estão distribuídas ao longo do documento e concluir que não existem regiões do texto com espaços ou distâncias consideráveis entre as sentenças com estrutura métrica. A maior distância encontrada entre as sentenças com estrutura métricas foi de 33 enquanto a média das distâncias de 3,4 sentenças.

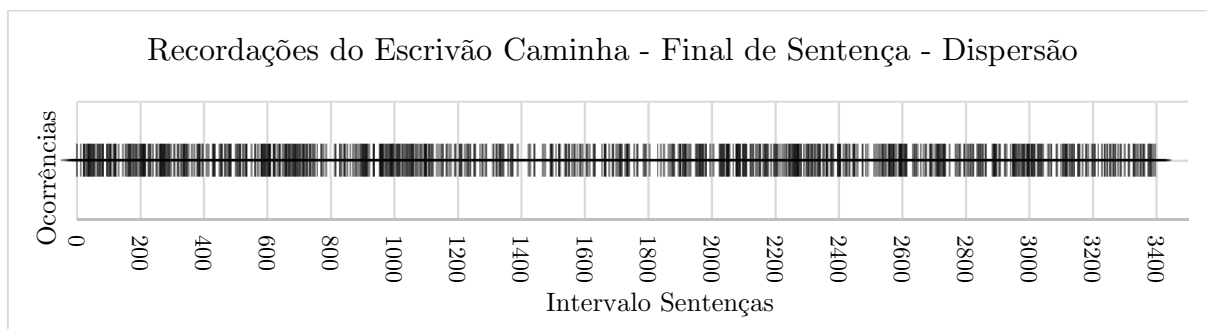


Figura 54: Recordações do Escrivão Caminha – Final de Sentenças - Distribuição das sentenças métricas ao longo do texto

Com relação aos padrões rítmicos característicos da poesia presentes nos finais de sentenças com estrutura métrica, dentre um total de 609 variações de posicionamento de tônicas, destacamos a presença de 16 heroicos, 11 martelos, 5 gaitas galega, 1 sáfico e 22 alexandrinos. A Tabela 39 apresenta o ranking das 10 variações de posicionamento de tônicas encontrados com maior frequência nesta etapa do experimento, enquanto o gráfico da Figura 55 o comportamento geral destas frequências.

Tabela 39: Recordações do Escrivão Caminha - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentenças

Posicionamento das tônicas	Frequência
2 6 10	12
1 4 8 11	12
1 4 7 10	11
3 7 10	11
3 6 10	11
2 5 8 10	10
2 6 11	10
3 6 11	10
3 6 8 10	9
3 5 8 10	9

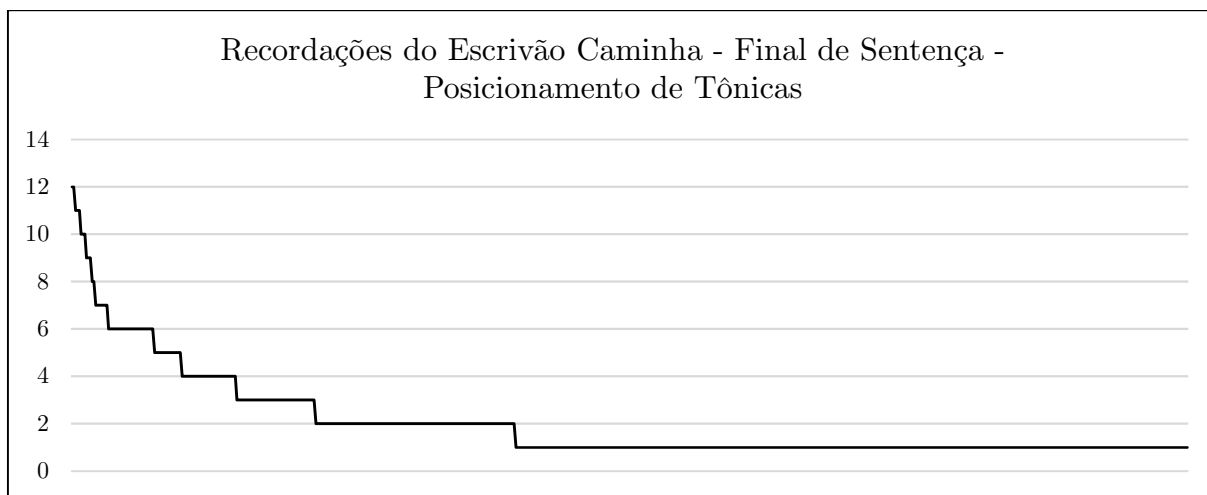


Figura 55: Recordações do Escrivão Caminha – Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Nesta obra de Lima Barreto, o MIVES conseguiu identificar inúmeras estruturas de versificação na prosa analisada. Além destas questões analisamos o comportamento relacionado a frequência absoluta de cada tipo encontrado em relação ao total de segmentos analisados e como estas estruturas estão distribuídas ao longo do texto. A Tabela 40 apresenta os resultados.

Tabela 40: Resumo das Experiências com Recordações do Escrivão Caminha.

Parâmetros	Resultado
Total de Sentenças	3202
Sentenças completas com estrutura métrica (total e percentual)	301 9,40%

Início de sentenças com estrutura métrica (total e percentual)	768 23,99%
Finais de sentenças com estrutura métrica (total e percentual)	755 23,58%
Quantidade de Decassílabos (sentenças, trechos iniciais e finais)	149 400 314 Total: 863
Quantidade de Dodecassílabos (sentenças, trechos iniciais e finais)	146 334 426 Total: 906
Distância entre sentenças (média, desvio padrão, mínimo e máximo)	10,24±13,78 (0-95) 3,41±4,15 (0-29) 3,50±4,22 (0-33)

Os valores relacionados a distância entre sentenças, de média e desvio padrão mostram que os valores de distância entre sentenças com estrutura métrica não são constantes, ou seja, não existe um padrão para ocorrência destas estruturas ao longo do texto. Outro ponto que merece destaque é a presença de 144 sentenças com estrutura métrica que obedecem a padrões rítmicos característicos da poesia (heroicos, martelos, sáficos, gaita galega e alexandrinos). Nesta obra de Lima Barreto o MIVES conseguiu apresentar um número relevante de segmentos com estruturas métricas. Não se pode tratar aqui questões de taxa de identificação ou precisão do processo de busca, por não termos fontes que nos permitam comparar com os resultados alcançados pelo MIVES, mas os resultados apresentados são capazes de fornecer indícios da presença de estruturas métricas característicos da poesia sobre esta obra de Lima Barreto.

4.4.3 Machado de Assis - Dom Casmurro.

Escrito por Machado de Assis e publicado em 1899 o romance tem seu lugar de destaque na literatura brasileira. O enredo gira em torno do personagem Bento Santiago, também narrador da história, constituída por relatos da sua própria vida, da mocidade a idade adulta, fazendo o paralelo entre os dois momentos, relatando dentre outras coisas sua experiência no seminário e seu caso com Capitu. Seguindo a metodologia aplicada nas experiências anteriores para realizar a busca por estruturas de versificação submetemos Dom Casmurro ao MIVES.

Concluído o processamento, o MIVES retornou 392 estruturas de versificação constituídas por sentenças completas. Sendo 183 decassílabos, 209 hendecassílabos e 224 dodecassílabos. A Tabela 41 apresenta alguns exemplos de estruturas de versificação encontradas pelo MIVES neste experimento.

Tabela 41: Exemplo de Sentenças Completas - Dom Casmurro

Sentença Completas com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
O mais é também análogo e parecido.	O/ mais / é/ tam/ bém / a/ ná /lo/go e/ pa/re/ ci /do.	12
Em verdade, pouco apareço e menos falo.	Em/ ver/ da /de,/ pou /co a/pa/re/ço e/ me /nos/ fa /lo.	12
Quando lhe disse o que era, fez-se cor de cera.	Quan /do/ lhe/ di /sse o/quee /ra,/ fez -/se/ cor / de/ ce /ra.	12
Cerrava os dentes, abanava a cabeça...	Ce/ rra /va os/ den /tes,/ a/ba/ na /va a/ca/ be /ça...	11
	Ce/ rra /va/ os/ den /tes,/ a/ba/ na /va a/ca/ be /ça...	12
Fui devagar, Mas ou o pé ou o espelho traiu-me.	Fui / de/va/ gar ,/ Mas/ ou o/ pé / ou o es/ pe /lho/ tra/ iu -/me.	12
O mais é também análogo e parecido.	O/ mais / é/ tam/ bém / a/ ná /lo/go e/ pa/re/ ci /do.	12

Seguindo a mesma metodologia de análise dos experimentos anteriores, no gráfico da Figura 56 é possível analisar a frequência absoluta das sentenças com estrutura métrica identificadas. As variações apresentadas pelo gráfico mostram regiões com maior densidade de sentenças com estrutura métrica, mas sem grandes amplitudes nos valores de frequência. Nas regiões do gráfico onde os valores apresentam-se de forma constante evidenciam intervalos onde não existe ocorrência de sentenças com estrutura métrica. Essa conclusão fica mais evidente quando analisamos a dispersão destas estruturas apresentadas pelo gráfico da Figura 57.

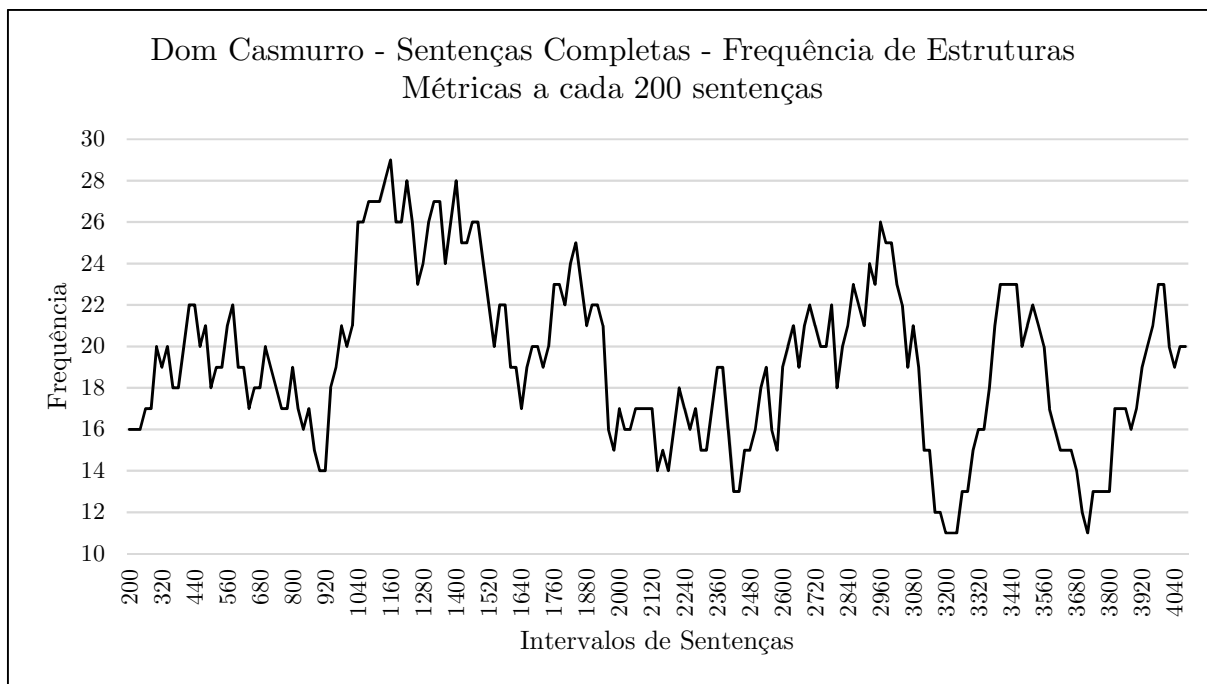


Figura 56: Dom Casmurro – Sentenças Completas - Distribuição da Frequências Absolutas.

No gráfico da Figura 57 é possível visualizar inúmeras regiões do texto em que não existem a ocorrência de sentenças com estrutura métrica. A médias das distâncias entre as estruturas métricas é de 9,34 enquanto da maior distância foi de 57 estruturas.

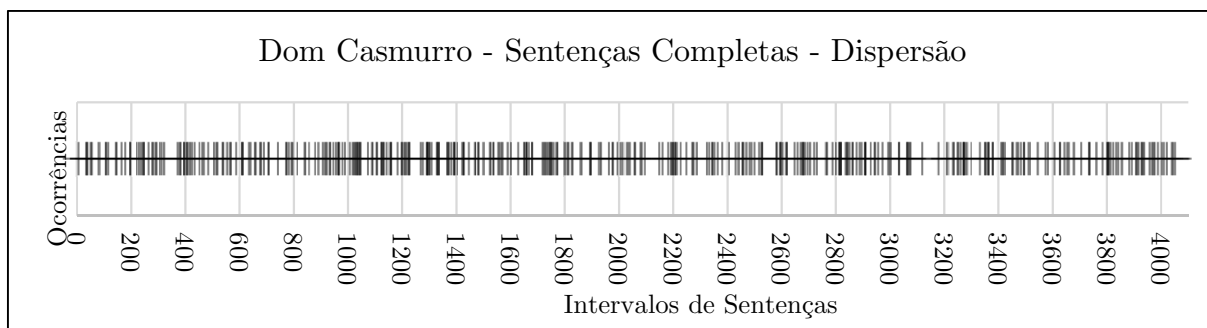


Figura 57: Dom Casmurro – Sentenças Completas - Distribuição das sentenças métricas ao longo do texto

Ampliando nossas análises sobre os resultados com Dom Casmurro, também analisamos os padrões rítmicos característicos da poesia, neste ponto o MIVES reportou 5 heroicos, 3 martelos e 6 alexandrinos. Estes padrões foram identificados dentre uma população de 400 variações de distribuição de tônicas, distribuídos entre as sentenças com estrutura métrica reportada pelo MIVES. A Tabela 42 apresenta um ranking dos 10 padrões rítmicos com maior frequência. O gráfico da Figura 58 mostra o comportamento relacionado a frequência com que os padrões ritmos ocorreram nas sentenças completas com estrutura métrica.

Tabela 42: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente – Sentenças Completas

Posicionamento de Tônicas	Frequência
1 4 6 10	10
1 5 7 11	7

2 5 8 11	6
2 5 8 10	6
3 7 10	6
3 5 7 10	6
2 4 6 10	6
1 3 5 10	6
2 5 7 10	5
1 4 7 10	5

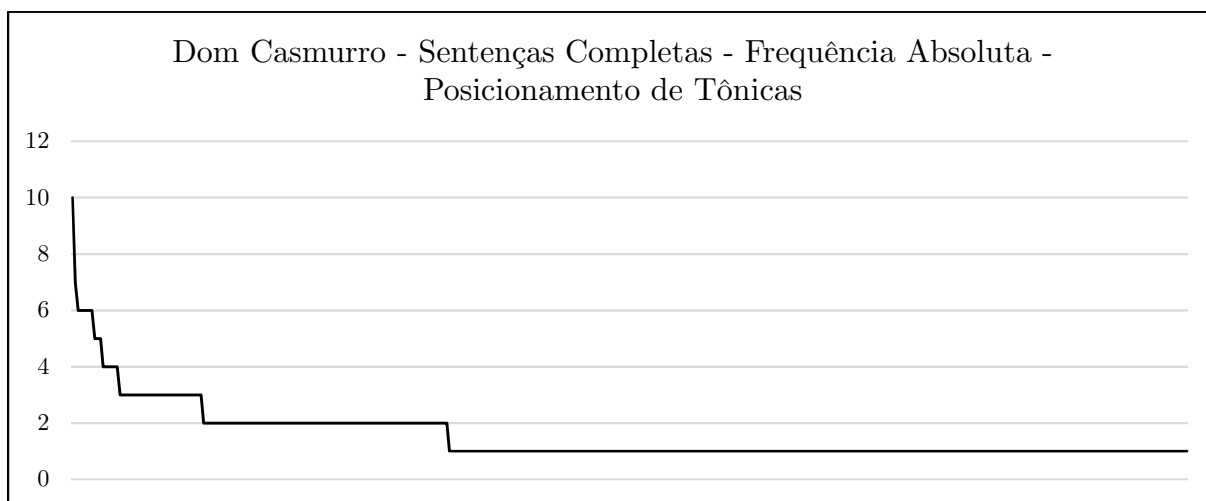


Figura 58: Dom Casmurro – Sentenças Completas- Frequência Absoluta - Posicionamento das Tônicas

Seguindo a trajetória do experimento, na segunda etapa, buscando por estruturas de versificação constituídas nos inícios de sentenças. Aqui o sistema identificou 941 sendo 500 decassílabos, 597 hendecassílabos e 453 dodecassílabos. A Tabela 43 apresenta alguns exemplos destas estruturas.

Tabela 43: Exemplo de Início de Sentenças - Dom Casmurro

Início de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
Cumprimentou-me, sentou-se ao pé de mim,	Cum/pri/men/ tou- /me,/ sen/ tou- /se ao/ pé / de/ mim ,	11
	Cum/pri/men/ tou- /me,/ sen/ tou- /se/ ao/ pé / de/ mim ,	12
Talvez a narração me desse a ilusão,	Tal/ vez / a/ na/rra/ ção / me/ de /sse a i/lu/ são ,	11
	Tal/ vez / a/ na/rra/ ção / me/ de /sse/ a i/lu/ são ,	12

Deste modo, viverei o que vivi,	Des/te/mo/do,/ vi/ve/rei o/ que/ vi/vi,	10
	Des/te/mo/do,/ vi/ve/re/i o/ que/ vi/vi,	11
Tudo se teria passado sem mais nada,	Tu/do/ se/ te/ria/ pa/ssa/do/ sem/mais/na/da,	11
	Tu/do/ se/ te/ri/a/ pa/ssa/do/ sem/mais/na/da,	12
E as vozes repetiam-se confusas:	E as vo/zes/ re/pe/ti/am-/se/ con/fu/sas:	10
	E/ as vo/zes/ re/pe/ti/am-/se/ con/fu/sas:	11
Dei um pulo, e antes que ela raspasse o muro,	Dei/ um/pu/lo,ean /tes/que e/la/ ras/pa/sse o/mu/ro,	11
	Dei/ um/pu/lo,/e an/tes/que e/la/ ras/pa/sse o/mu/ro,	12

A frequência absoluta relacionada a distribuição dos inícios de sentença com estrutura métrica pode ser visualizado pelo gráfico da Figura 59. Neste gráfico é possível visualizar que existem variações nas distribuições de frequência das sentenças com estrutura métrica identificadas nesta etapa. A região com menor densidade possui 32 enquanto de maior 58 sentenças com estrutura métrica.

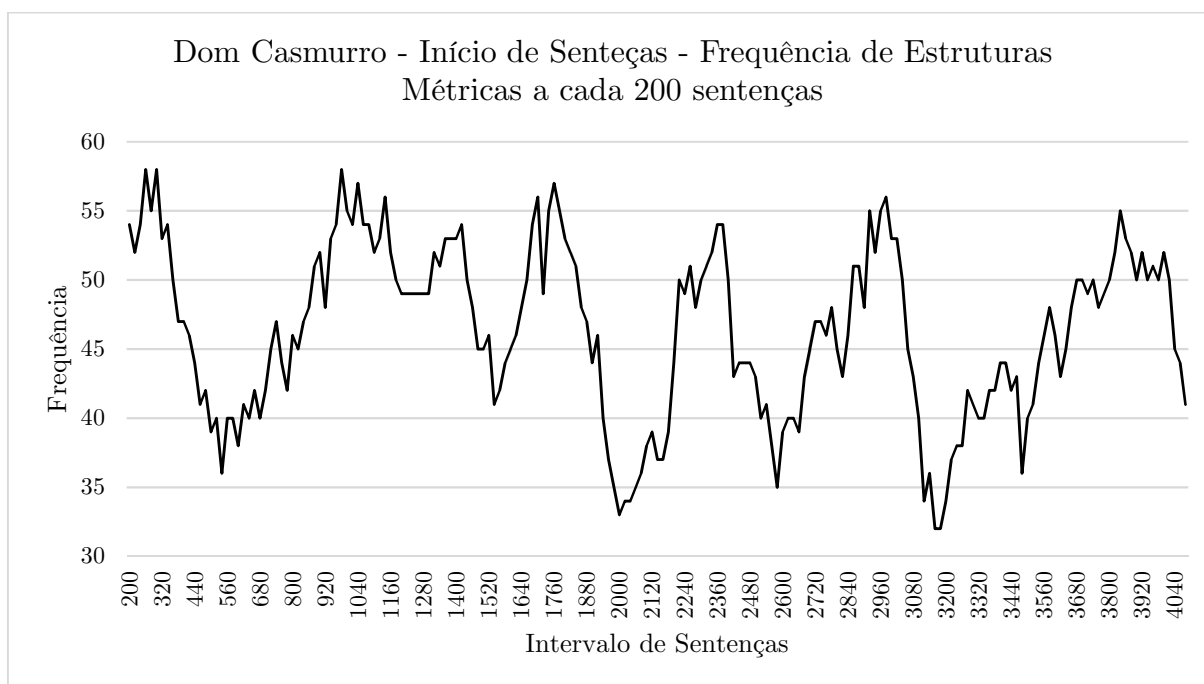


Figura 59: Distribuição de frequências absolutas – Dom Casmurro – Início de Sentenças

O comportamento relacionado a dispersão das sentenças com estrutura métrica nos inícios das sentenças pode ser visualizado no gráfico da Figura 60. Neste gráfico é possível ver que, embora exista uma distribuição uniforme, as sentenças com estrutura métrica estão distribuídas ao longo de todo o texto.

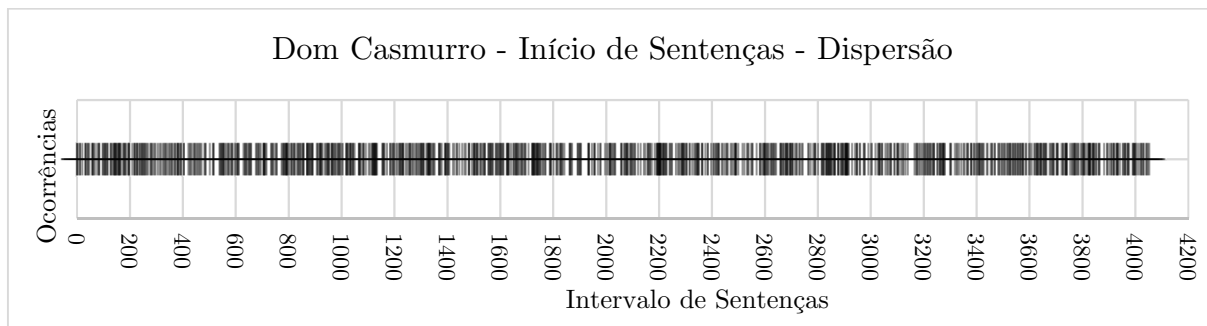


Figura 60: Dom Casmurro – Início de Sentenças - Distribuição das sentenças métricas ao longo do texto

Para completar nosso experimento analisamos qual o comportamento relacionado a frequência da distribuição das vogais tônicas sobre os inícios de sentenças com estrutura métrica apresentados pelo MIVES. O resultado pode ser observado no gráfico da Figura 61.

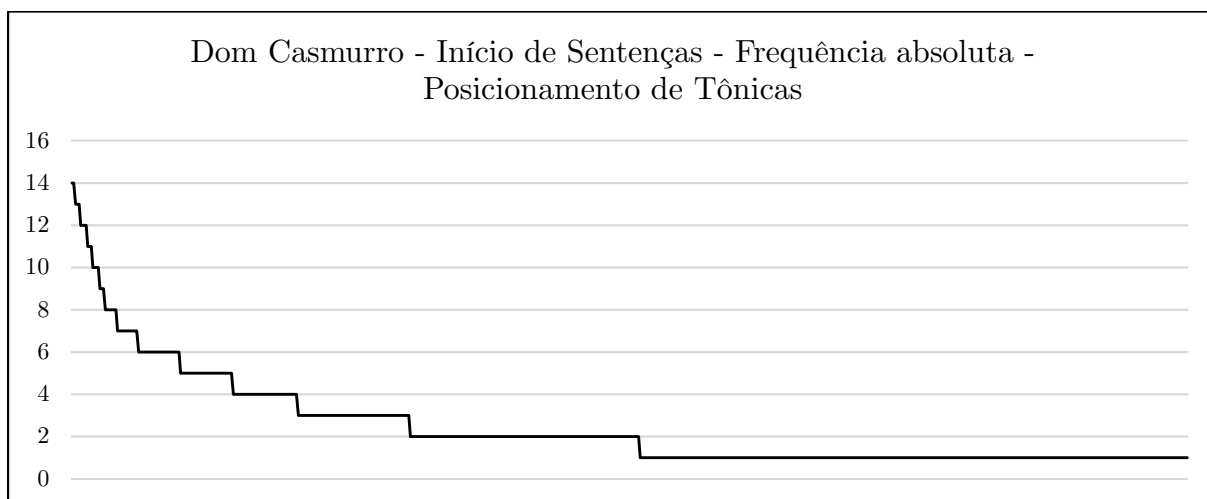


Figura 61: Dom Casmurro – Início de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Além do gráfico da Figura 61, a Tabela 44 possui um ranking das 10 distribuições de tônicas que ocorreram com maior frequência dentre as 637 variações de posicionamento reportadas pelo MIVES. Neste universo, os padrões rítmicos característicos da poesia, destacam-se os 14 heroicos, 13 martelos, 3 gaíta galega, 24 alexandrinos.

Tabela 44: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente – Início de Sentenças

Posicionamento de Tônicas	Frequência
2 5 7 10	14
1 4 6 10	14
3 7 10	13
3 5 7 10	13
3 6 10	13

1 4 7 10	12
2 5 8 10	12
2 4 6 10	12
1 3 5 10	12
2 5 8 11	11

Na última etapa buscamos por finais de sentenças com estrutura métrica. De um total de 997 estruturas de versificação distribuídas pelo texto entre 387 decassílabos, 619 hendecassílabos e 598 dodecassílabos. A Tabela 45 apresenta alguns exemplos dos achados desta etapa do experimento com Dom Casmurro.

Tabela 45: Exemplo de Final de Sentenças - Dom Casmurro

Final de Sentença com Estrutura Métrica	Escansão segundo MIVES	Nº Sílabas Poéticas
deram curso à alcunha, que afinal pegou.	de /ram/ cur /so à al/ cu /nha,/ que a/fi/ nal /pe/ gou .	11
	de /ram/ cur /so/ à al/ cu /nha,/ que a/fi/ nal /pe/ gou .	12
para atribuir-me fumos de fidalgo.	pa /ra a/tri/bu/ ir -/me/ fu /mos/ de/fi/ dal /go.	11
	pa /ra/ a/tri/bu/ ir -/me/ fu /mos/ de/fi/ dal /go.	12
varanda ao fundo, as mesmas alcovas e salas.	va/ ran /da ao/ fun /do, as/ mes /mas/al/ co /vas/ e/ sa /las.	12
de uma elegância pobre e modesta.	de/ u /ma e/le/ gân /cia/ po /bre e/ mo/ des /ta.	10
	de/ u /ma/ e/le/ gân /cia/ po /bre e/ mo/ des /ta.	11
	de/ u /ma/ e/le/ gân /cia/ po /bre/ e/ mo/ des /ta.	12
gostava do elogio; era um elogio.	gos/ ta /va/ do e/lo/ gi /o;/ e/ra um/ e/lo/ gi /o.	12
assim o mestre de	a/ ssim / o/ mes /tre/ de/ pri/ mei /ras/ le /tras.	10

primeiras letras.	a/ ssim / o/ mes /tre/ de/ pri/ me /i/ras/ le /tras.	11
-------------------	---	----

Em Dom Casmurro, pelos números apresentados, fica evidente a presença de estruturas com metros superiores, entre 10 e 12 sílabas poéticas. Para visualizar o comportamento de como as estruturas ocorrem ao longo do texto basta analisar o gráfico da Figura 62. Ao analisar este gráfico é possível concluir que, embora existam regiões de maior concentração, as sentenças estão distribuídas ao longo de todo o texto. Esta conclusão se reforça ao analisar o gráfico de dispersão da Figura 63, nele é demonstrado a distribuição destas sentenças ao longo do texto. Os finais de sentença aconteceram em Dom Casmurro de forma distribuída ao longo de todo o texto, embora exista uma natural concentração deste tipo de estrutura em alguns intervalos de sentenças.

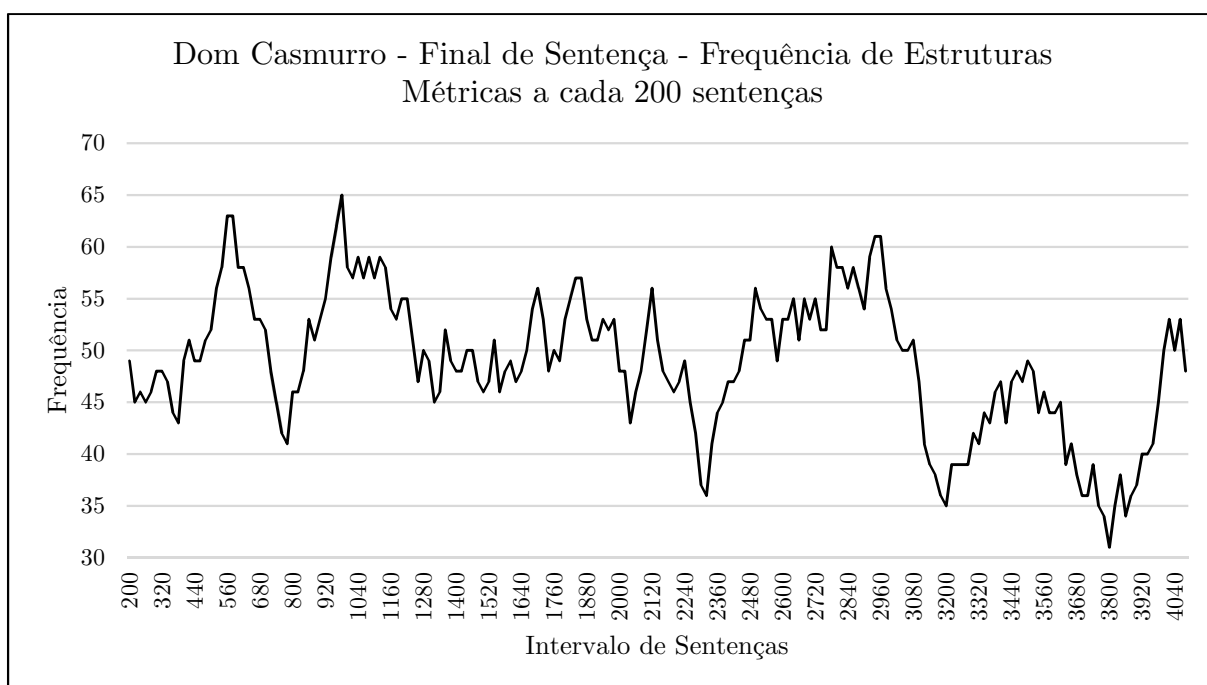


Figura 62: Dom Casmurro – Final de Sentença - Distribuição da Frequências Absolutas.

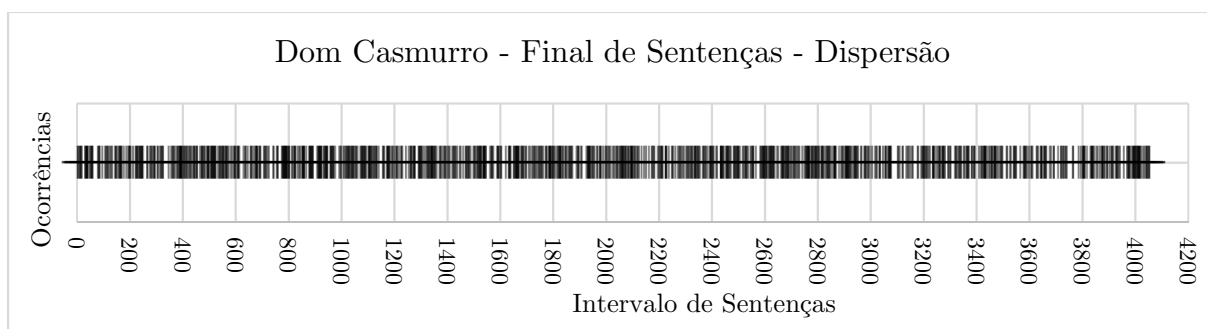


Figura 63: Dom Casmurro – Final de Sentenças - Distribuição das sentenças Métricas ao Longo do Texto

Para encerrar as experiências com Dom Casmurro analisamos o posicionamento das

tônicas sobre os finais de sentenças reportada pelo MIVES. Um ranking com as 10 variações de maior frequência, extraídos de uma população de 708 variações, pode ser observado na Tabela 46. Além destes, os padrões rítmicos característicos da poesia identificados são constituídos por 11 heroicos, 11 martelos, 4 gaítas galega, 2 sáficos e 26 alexandrinos. A frequência de como estes padrões ocorrem dentre os finais de sentença com estrutura métrica pode ser observado no gráfico da Figura 64.

Tabela 46: Dom Casmurro - Ranking dos 10 posicionamentos de tônicas mais frequente – Final de Sentenças

Posicionamento de Tônicas	Frequência
2 5 8 11	19
2 4 7 11	16
2 5 9 12	14
3 6 8 11	13
1 4 6 10	12
1 5 7 11	12
2 4 7 12	12
1 4 7 11	11
1 4 7 10 12	11
2 4 7 9 12	11

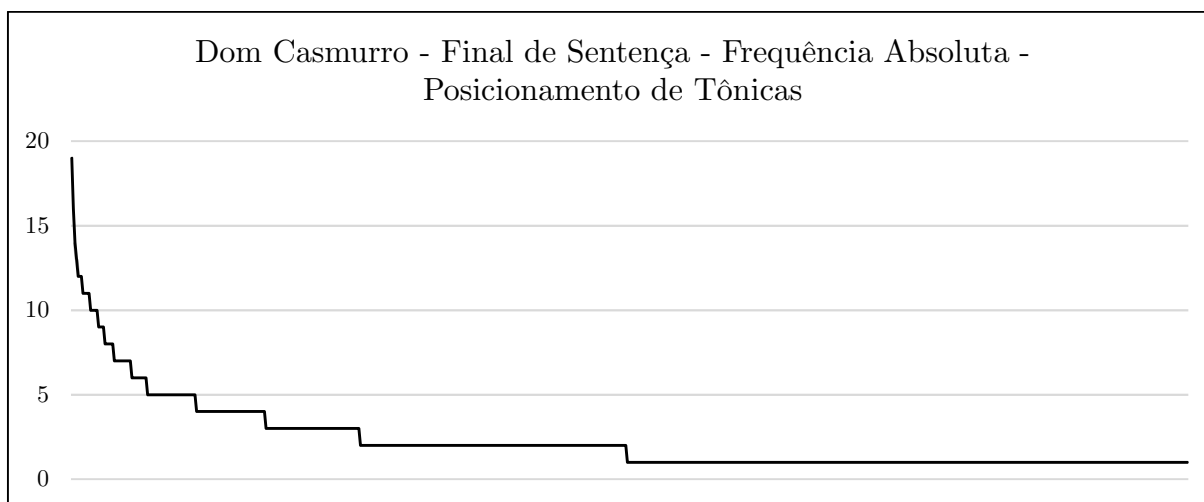


Figura 64: Dom Casmurro – Final de Sentenças - Frequência Absoluta - Posicionamento das Tônicas

Tabela 47: Resumo das Experiências com Dom Casmurro

Parâmetros	Resultado
Total de Sentenças	4010
Sentenças completas com estrutura métrica (total e percentual)	392 9,78%

Início de sentenças com estrutura métrica (total e percentual)	941 23,47%
Finais de sentenças com estrutura métrica (total e percentual)	997 24,86%
Quantidade de Decassílabos (sentenças, trechos iniciais e finais)	183 500 387 Total: 1070
Quantidade de Dodecassílabos (sentenças, trechos iniciais e finais)	224 453 598 Total: 1275
Distância entre sentenças (média, desvio padrão, mínimo e máximo)	9,34±9,55 (0-57) 3,31±3,83 (0-25) 3,07±3,62 (0-25)

Os valores apresentados pela Tabela 47, mostram a ocorrência de numerosas estruturas de verificação em Dom Casmurro. Com maior concentração de estruturas dodecassilábicas de finais de sentença. Os altos valores de desvio padrão também revelam que não existe uma regularidade de distâncias entre os segmentos identificados. Dentre os padrões rítmicos característicos da poesia, encontramos cerca de 124 ocorrências entre heroicos, martelos, gaitas galega e alexandrinos. Os dados apresentados nestes experimentos são capazes de mostrar que existem algumas regiões do texto com métrica ou padrão rítmico característicos da poesia no Brasil.

4.4.4 Considerações a Respeito dos Experimentos

Em todos os experimentos desta seção encontramos por estruturas de versificação (sentenças com estrutura métrica) com métrica entre 10 e 12 sílabas poéticas. Nosso objetivo foi alcançado por termos revelado sentenças em variados pontos dos textos analisados com estrutura métrica. O MIVES forneceu, sempre em que foi possível realizar a escansão, diferentes formações escandidas para o mesmo segmento anunciado com estrutura métrica. As informações extraídas como, os segmentos escandidos, as estatísticas de quantidade de versos por tipo encontrado, e a visualização de como estão distribuídas estas sentenças ao longo do texto, bem como os valores de distância e desvio padrão, e os padrões rítmicos característicos da poesia, formam o conjunto de informações que se fizeram necessárias para as primeiras análises realizadas neste trabalho.

A grande questão é se os resultados apresentados pelo MIVES são, ou podem servir como indícios, capazes de motivar uma pesquisa mais aprofundada sobre a presença de aspectos versificatórios sobre as obras analisadas nos três últimos experimentos. Diante dos números apresentados e das informações estruturais das sentenças indentificadas com estrutura métrica, fica claro a existência do fenômeno relacionado a presença de estruturas de versificação sobre textos da prosa literária brasileira. Os indícios existem, mas cabe ao usuário do MIVES, baseados nos resultados apresentados pelo software, decidir sobre intertar ou não em uma pesquisa mais detalhada sobre o exemplar da prosa literária alvo do estudo.

O MIVES tem o principal propósito de escandir estruturas métricas subliminares na prosa literária em língua portuguesa. As estratégias utilizadas pelo MIVES e os resultados obtidos com as experiências aqui descritas, representam os passos iniciais em um campo de pesquisa ainda inexplorado sobre a prosa literária em língua portuguesa.

Capítulo 5

Considerações Finais

As pesquisas desenvolvidas durante a construção desta dissertação nos permitiram atingir objetivo de propor uma metodologia computacional capaz de realizar a identificação automática de estruturas de versificação da prosa literária brasileira. Para tornar concreto esse objetivo implementamos o MIVES, uma ferramenta computacional capaz de realizar a busca por estruturas de versificação ou sentenças com estrutura métrica em textos da prosa literária brasileira, um processo que envolve a manipulação de estruturas de texto em prosa (e.g determinação dos limites da estrutura métrica, separação silábica, transformação da sílaba gramatical em sílaba poética, metrificação e classificação) com o intuito de identificar aspectos característicos da poesia. Além disso desenvolvemos uma interface de análise que permite ao usuário visualizar os resultados do processo de busca realizado pelo MIVES. Validamos o processo de escansão submetendo ao MIVES um texto poético com PPPP versos de métrica conhecida e atingimos uma precisão de mais de 94%. Exploramos Os Sertões e, além de alcançar resultados já anunciados por Almeida[1946] e Campos[1996], apresentamos inúmeros dados não apresentados nos artigos dos percussores que serviram como inspiração para realização deste trabalho (eg. como se dá a distribuição destas estruturas ao longo do texto, com que frequência estas ocorrem, as classificações específicas de cada verso encontrado, seu padrão rítmico). Para expandir nossas pesquisas submetemos ao MIVES outras obras de Euclides da Cunha e revelamos também que nestas existem elementos poéticos similares a aqueles encontrados em Os Sertões (eg. inúmeras estruturas deca e dodecassilábicas com diferentes padrões rítmicos). Comparamos os resultados obtidos e chegamos a conclusão de que existe um padrão no comportamento da escrita de Euclides que tende a construção de estruturas deca e dodecassilábicas.

Fomos além de algumas obras de Euclides de Cunha, submetemos ao MIVES outras obras clássicas da prosa brasileira a saber: Canaão, Recordações do Escrivão Caminha e Dom Casmurro. Nas três obras encontramos e apresentamos a presença de estruturas com métrica poética e, da mesma forma que fizemos com as obras de Euclides da Cunha, apresentamos dados de como se dá o comportamento versificatório destes autores nos textos analisados. Não podemos afirmar que este é um comportamento comum nos textos literários daquela época, nosso objetivo é de apenas trazer a luz informações, antes ocultas, e viabilizar que estas sejam analisadas por um especialista da área.

Por não termos encontrados trabalhos similares nesta pesquisa, mesmo que para outros

idiomas, afirmamos que nossa pesquisa é o primeiro passo em um terreno ainda vasto de possibilidades de pesquisa sobre a prosa literária brasileira. O MIVES, permite que análises que levariam dias ou até meses para serem realizadas possam ser feitas em poucos minutos. A velocidade com que o MIVES pode trazer a luz estruturas versificação, antes ofuscadas pelo enredo da obra ou ocultadas pela estrutura do texto, poderá tornar possível a criação de um corpus de textos da literatura com informações sobre obras da literatura brasileira, antes inimagináveis pelo custo e tempo que levariam para ser construídos.

Referências Bibliográficas

- [Ali 2006] Ali, S. (2006) *Versificação Portuguesa*. EDUSP.
- [Almeida 1946] Almeida, G. (1946). *A poesia d'Os Sertões*. Diário de São Paulo.
- [Araújo e Mamede (2002)] Araújo, P. e Mamede, N. (2002). *Classificador de poemas*. Lisboa, Portugal.
- [Bernucci e Hardman 2009] Bernucci, M. e Hardman, F. (2009) *Poesia reunida / Euclides da Cunha; organização, estabelecimento de textos, introduções, notas e índices*. São Paulo: Unesp.
- [Brandão 2002] Brandão, A., *Os Sertões uma revolução literária*. In Os Sertões Campanha de Canudos. São Paulo: Martin Claret.
- [Campos 1996] Campos, A. (1996). *Transertões*. Folha de São Paulo.
- [Cunha e Cintra 2008] Cunha, C. e Cintra, L. (2008) *Nova Gramática do Português Contemporâneo*. 5. ed. Rio de Janeiro: Lexikon.
- [Gervas 2000] Gervas, P. (2000) *A logic programming application for the analysis of Spanish verse*. In: Computational Logic. Springer Berlin Heidelberg. p. 1330-1344.
- [Goldstein 2006] Goldstein, N., *Versos, sons, ritmos*. 14. ed. São Paulo: Ática.
- [Jakobson 1985] Jakobson, R. (1988). *Dialogues*. MIT Press.
- [Mittmann 2016] Mittmann, A., Santos, A. L. e Wangenheim, A. Aoidos: A System for the Automatic Scansion of Poetry Written in Portuguese. In: International Conference on Intelligent Text Processing and Computational Linguistics.
- [Moisés 2002] Moisés, M. (2002) *Dicionário de Termos Literários*. Editora Cultrix.
- [Neto et al. 2015] Neto, N., Rocha, W. e Sousa, G. (2015) *An open-source rule-based syllabification tool for brazilian portuguese*. Journal of the Brazilian Computer Society. Springer, p. 1–10.
- [Niskier 2009] Niskier, A. (2003) *Estilo e escrita*. Disponível em: <<http://www.academia.org-br/artigos/escrita-e-estilo>> Acesso em maio de 2017
- [Silva 2014] Silva, P. (2014). *Teoria da Literatura*. São Paulo: Pearson Education do Brasil.
- [Spina 2003] Spina, S. (2003). *Manual de Versificação Românica Medieval*. Ateliê Editorial.
- Williams, A. C., Carroll, H. D., Wallin, J. F., Brusuelas, J., Fortson, L., Lamblin, A. F., & Yu, H. (2014, October). *Identification of ancient greek papyrus fragments using genetic sequence alignment algorithms*. In e-Science (e-Science), 2014 IEEE

10th International Conference on (Vol. 2, pp. 5-10). IEEE.

Anexo 01

Parâmetros de escansão utilizados durante os experimentos.

As combinações de vogais apresentadas pelo quadro a seguir representam para o MIVES os possíveis encontros entre vogais que podem ser tratados pelo MIVES como possíveis casos de fenômenos fonológicos.

àa	àè	áá	ââ	ãë	ài	aã	ãõ	èâ	éá	éõ	êâ	êã	îâ	îà	íõ	ïî	ïè	ðè	òâ	oà	óõ	õõ	õî
àe	àé	âi	âã	ãî	ãí	aä	aò	èã	éâ	eí	êã	ëã	îã	îä	ìò	ïí	ïè	ðé	òã	óã	oò	òò	uo
ài	àê	áé	âè	ãî	ãí	âè	ãõ	èä	éã	íu	êã	ëã	îä	îá	íõ	ïî	óo	ðê	òä	oá	óõ	óó	õî
ào	àë	ai	âé	ãî	ãï	aã	aó	èå	eà	eí	îi	ëå	ðî	îå	íó	ïî	ïî	òo	òå	óå	oó	ôó	õõ
àu	äo	áé	âê	ãî	ãð	ei	aô	ìi	éä	éò	êå	ïi	îå	îä	îò	îð	ïî	ðê	öi	oâ	oô	ôô	ùu
áa	àì	áé	âë	ãð	èu	áé	aõ	èè	eá	eï	èè	èè	ìè	îã	îõ	òu	ïî	ðî	òè	oã	oõ	ôõ	õò
áe	àí	áé	âì	ãò	ãò	âè	ãù	èé	ie	éó	éè	ëé	ìé	îä	îù	îò	oo	ðí	òé	oä	óù	òù	õó
ái	âî	ão	âí	ãò	ãó	ãè	aõ	èè	íi	eð	êè	ëè	ìè	ìè	îò	îó	ïî	ðî	òè	óè	oö	óù	õò
áo	àï	ái	âî	ãò	ãò	éó	ãú	èè	eã	éó	íó	ëè	ðó	îå	íú	îò	îð	ðî	òè	oå	óú	óù	uu
áu	àð	ái	âï	ãò	ãò	aè	ãù	ìo	eâ	iu	êè	ïo	ìè	ìé	íú	îð	óu	ðð	öo	óé	óù	ôù	õõ
âa	äu	ái	âð	ãò	ãò	ài	ãü	èi	eã	éõ	èi	èi	ìi	ìé	òa	îò	îò	òu	òi	óé	óü	úa	õõ
âe	àò	ao	âò	ãù	ãù	aé	êa	èi	eä	eò	éí	éí	ìí	ìè	íú	îù	îó	ðò	òí	óè	óù	òä	õù
âi	àó	ái	âó	ãù	ãù	ái	aù	èí	éè	éö	êí	ëí	îí	ìè	íù	îú	îò	ðó	òí	óè	óú	ua	õú
âo	àõ	ãð	âò	ãù	ãù	aé	aú	èí	eã	eó	êí	ëí	îí	îí	íú	îù	ou	ðò	òí	óí	óú	õá	õù
âu	àõ	ãu	ãõ	èa	ãü	ãí	aú	èð	ii	eô	êð	ëð	ìð	ìé	íú	îù	îò	ðò	òð	óé	òà	ùe	õü
ãa	àö	áó	âò	ãü	éa	aë	aü	ìu	éé	eõ	îu	ïu	ðu	íí	îà	óa	îö	ðò	öu	óí	óü	õâ	úa
ãe	àù	áó	âù	ãà	ãà	eo	êe	èò	éè	éù	èò	èò	ìò	ìè	íü	îà	îù	ðù	òò	óè	óá	õã	òà
ãí	áú	áó	âú	èe	ea	ãí	èi	èó	íó	eö	éó	éó	ìó	îí	òe	oa	íú	ðú	òó	óí	óâ	õã	óá
ão	áü	au	âù	ãá	ãá	ai	èo	èò	éè	éú	èò	èò	ìò	ìè	îá	îá	îù	ðù	òò	óè	óã	ue	úe
ãü	áü	áó	âü	ãä	ée	ãð	èu	èõ	eè	éú	êõ	ëõ	ìõ	îí	îá	óe	ôa	ðü	òõ	óí	óä	õã	õä
äa	ää	áó	ãà	ãã	ãã	ái	èa	èò	éí	îa	êò	ëò	ìò	ìi	îã	îã	íü	ôa	òò	óì	óå	ùí	õã
àà	áà	áú	ãá	ãä	ãã	éu	èe	èù	eé	éü	èù	èù	ìù	íð	îã	îã	ðà	ðe	òù	óð	òè	èè	òä
áá	aa	áú	ãâ	èi	aà	ái	èi	èú	éí	eù	éú	éú	íú	íí	òi	îã	òe	öi	òú	óí	óé	ui	õä

