

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

MARDEN ALEANDRO RANGEL

**PROPOSTA METODOLÓGICA PARA ANÁLISES DE OBRAS LITERÁRIAS
ATRAVÉS DA LINGUÍSTICA DE *CORPUS*: O CASO DE PAULO COELHO**

Uberlândia

2022

MARDEN ALEANDRO RANGEL

**PROPOSTA METODOLÓGICA PARA ANÁLISES DE OBRAS LITERÁRIAS
ATRAVÉS DA LINGUÍSTICA DE *CORPUS*: O CASO DE PAULO COELHO**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Letras e Linguística, da Universidade Federal de Uberlândia, como requisito parcial à obtenção do título de Mestre em Estudos Linguísticos.

Área de concentração: Estudos em Linguística e Linguística Aplicada.

Linha de pesquisa 1: Teoria, descrição e análise linguística.

Orientador: Prof. Dr. Guilherme Fromm

Uberlândia

2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

R196 2022	<p>Rangel, Marden Aleandro, 1973- Proposta metodológica para análises de obras literárias através da Linguística de Corpus: o caso de Paulo Coelho [recurso eletrônico] / Marden Aleandro Rangel. - 2022.</p> <p>Orientador: Guilherme Fromm. Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Estudos Linguísticos. Modo de acesso: Internet. Disponível em: http://doi.org/10.14393/ufu.di.2022.148 Inclui bibliografia. Inclui ilustrações.</p> <p>1. Linguística. I. Fromm, Guilherme, 1968-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Estudos Linguísticos. III. Título.</p> <p style="text-align: right;">CDU: 801</p>
--------------	--

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
 Coordenação do Programa de Pós-Graduação em Estudos Linguísticos
 Av. João Naves de Ávila, nº 2121, Bloco 1G, Sala 1G256 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
 Telefone: (34) 3239-4102/4355 - www.ileel.ufu.br/ppgel - secpffel@ileel.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Estudos Linguísticos				
Defesa de:	Dissertação de Mestrado Acadêmico				
	15 de fevereiro de 2022	Hora de início:	14:00	Hora de encerramento:	15:00
Matrícula do Discente:	11922ELI013				
Nome do Discente:	Marden Aleandro Rangel				
Título do Trabalho:	Proposta metodológica para análises de obras literárias através da Linguística de <i>Corpus</i> : o caso de Paulo Coelho				
Área de concentração:	Estudos em Linguística e Linguística Aplicada				
Linha de pesquisa:	Teoria, descrição e análise linguística				
Projeto de Pesquisa de vinculação:	Linguística de <i>Corpus</i> e Ciências do Léxico: compilação de <i>corpora</i> , descrição linguística e treinamento/desenvolvimento de <i>software</i>				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Estudos Linguísticos, assim composta: Professores Doutores: Celina Márcia de Souza Abbade - UNEB; Márcio Issamu Yamamoto - UFJ; Guilherme Fromm - PPGEL/UFU, orientador do candidato.

Iniciando os trabalhos o presidente da mesa, Dr. Guilherme Fromm, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por MÁRCIO ISSAMU YAMAMOTO, Usuário Externo, em 15/02/2022, às 15:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por Guilherme Fromm, Professor(a) do Magistério Superior, em 15/02/2022, às 16:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por CELINA MÁRCIA DE SOUZA ABBADE, Usuário Externo, em 15/02/2022, às 18:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 3377559 e o código CRC E7605813.

MARDEN ALEANDRO RANGEL

**PROPOSTA METODOLÓGICA PARA ANÁLISES DE OBRAS LITERÁRIAS
ATRAVÉS DA LINGUÍSTICA DE *CORPUS*: O CASO DE PAULO COELHO**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Letras e Linguística, da Universidade Federal de Uberlândia, como requisito parcial à obtenção do título de Mestre em Estudos Linguísticos.

Uberlândia, 27 de fevereiro de 2022.

Banca Examinadora

Prof. Dr. Guilherme Fromm – UFU
(orientador)

Profª. Dra. Celina Márcia de Souza Abbade (UNEB)

Prof. Dr. Márcio Issamu Yamamoto (UFJ)

DEDICATÓRIA

À minha mãe, Nelsa, e ao meu pai, Antônio, pela dignidade, dedicação incansável e apoio irrestrito que sempre tiveram com nossa família.

Ao meu esposo, Wendel, pelo amor e companheirismo em todos os momentos de nossa caminhada.

Esta conquista também é de vocês.

AGRADECIMENTOS

A Deus pela vida e por me guiar em toda a minha caminhada.

À minha mãe, Nelsa Maria Rangel, e ao meu pai, Antônio Aparecido Rangel, por dedicarem suas vidas à nossa família de forma incansável e sempre baseada no amor, honestidade, justiça e dignidade.

Ao meu esposo, Wendel Silva Nascimento, pela nossa vida de amor e companheirismo.

À minha irmã, Adriana Abadia Rangel Torres, ao meu cunhado, Valdes Barsanulfo Torres, e ao meu sobrinho, Vinícius Rangel Torres, pela torcida e alegria a cada conquista por mim obtida.

À minha sogra, Maria de Fátima Silva Nascimento, e ao meu sogro, Celso Vieira do Nascimento (*in memoriam*), por me aceitarem em sua família como se fosse um filho.

Ao meu orientador, Dr. Guilherme Fromm, um exemplo que eleva ao máximo as boas qualidades do que é ser professor e orientador, pela disposição, direcionamentos e saberes compartilhados.

Aos professores componentes da banca examinadora, Dra. Celina Márcia de Souza Abbade e Dr. Márcio Issamu Yamamoto, pela disponibilidade em ler meu trabalho, avaliar e fazer apontamentos, contribuindo com minha caminhada acadêmica.

Aos profissionais do Instituto de Letras e Linguística, da Universidade Federal de Uberlândia, especialmente às secretárias Maria Virgínia Dias de Ávila e Luana Alves da Silva, pela presteza e disponibilidade.

Aos colegas dos grupos de estudo, GPELC e PLEX, pelo companheirismo e incentivo, principalmente à Dra. Solange Aparecida Faria Cardoso pelo convite para conhecer a Linguística de *Corpus*, sendo este o início dos estudos na área.

A todos os meus professores, de todas as etapas escolares, que fizeram e continuam fazendo parte de minha formação acadêmica.

A todos os amigos, principalmente à Patrícia Aparecida Amaral e ao Fillipe Gomes de Sousa Alves, pela amizade, conversas acadêmicas e incentivos mútuos.

A todos que, de alguma forma, contribuíram para esta conquista.

Esta conquista é de todos nós.

RESUMO

Esta pesquisa objetivou desenvolver uma proposta metodológica capaz de analisar obras literárias através da Linguística de *Corpus*. Nosso principal intuito foi promover uma união entre a Linguística e a Literatura, por meio de ferramentas capazes de verificar aspectos lexicais e estilísticos de um ou mais autores e de suas obras. Utilizamos como pressupostos teóricos estudos sobre o Léxico, a Estilística, a Linguística de *Corpus* e a Estilística de *Corpus* e como metodologia a Linguística de *Corpus*. Entre nossos referenciais, constam, entre outros, Abbade (2011), Barbosa (1991), Berber Sardinha (1999, 2000, 2009), Biderman (1978, 1987, 2001), Finatto (2011), Fromm (2003), Mahlberg (2020), Martins (2012), Shepherd (2012, 2013, 2019) e Tagnin (2018). Nossos *corpora* de estudo foram os romances literários de Paulo Coelho, um dos autores mais lidos na atualidade, além de obras famosas de diversos escritores brasileiros, produzidas em diferentes épocas e diversas escolas literárias. A utilização de tais *corpora* possibilitou verificar o estilo lexical de Paulo Coelho, em aspectos como acessibilidade textual, complexidade textual, densidade lexical, palavras-chave, campos lexicais e uso de *hapax legomena*. Também comparamos a densidade lexical e a complexidade textual de *O alquimista*, de Paulo Coelho, estabelecendo paralelos com as obras dos demais autores. Na elaboração da proposta promovemos a coleta e montagem dos *corpora* que incluíram digitalizações, *downloads*, conversões, conferências e correções das obras. Em seguida, fizemos as análises através do programa *WordSmith Tools 6.0* (SCOTT, 2012), pelas ferramentas WordList, KeyWord e Concord; e pelo índice Flesch. Os resultados nos permitiram elaborar a proposta metodológica de análises de obras literárias através da Linguística de *Corpus*, além de analisar aspectos estilísticos e lexicais dos romances literários de Paulo Coelho.

Palavras-chave: Léxico. Estilística. Linguística de *Corpus*. Paulo Coelho.

ABSTRACT

This research aimed to develop a methodological proposal capable of analyzing literary works through Corpus Linguistics. Our main intention was to promote a union between Linguistics and Literature, using tools capable of verifying lexical and stylistic aspects of one or more authors and their works. We used as theoretical presuppositions studies on the Lexicon, Stylistics, Corpus Linguistics and Corpus Stylistics, having Corpus Linguistics as methodology. Among our references, we use Abbade (2011), Barbosa (1991), Berber Sardinha (1999, 2000, 2009), Biderman (1978, 1987, 2001), Finatto (2011), Fromm (2003), Mahlberg (2020), Martins (2012), Shepherd (2012, 2013, 2019) and Tagnin (2018). Our corpora of study were composed by Paulo Coelho's literary novels, one of the most read authors today, as well as famous works by several Brazilian writers, produced at different times and various literary schools. The use of such corpora made it possible to verify the lexical style of Paulo Coelho, in aspects such as textual accessibility, textual complexity, lexical diversity, keywords, lexical and semantic fields and the use of hapax legomena. We also compared the lexical density and textual complexity of *The Alchemist*, by Paulo Coelho, establishing parallels with the works of other authors. For the proposal, we collected and assembled the corpora that included text digitalization, downloads, conversions, conferences and corrections. Then, we performed the analysis using tools such as WordSmith Tools 6.0 program (SCOTT, 2012) by the WordList, KeyWord and Concord tools; and by the Flesch index. The results allowed us to elaborate the methodological proposal for the analysis of literary works through Corpus Linguistics, and also analysis of stylistic and lexical aspects of Paulo Coelho's literary novels.

Keywords: Lexicon. Stylistics. Corpus Linguistics. Paulo Coelho.

LISTA DE FIGURAS

Figura 1: Resultado da busca por “Linguística de <i>Corpus</i> ” pela Plataforma Lattes, do CNPq.....	41
Figura 2 – Resultado da busca por “Linguística de <i>Corpus</i> ” pelo Google Livros.....	41
Figura 3 – Resultado da busca por “Linguística de <i>Corpus</i> ” pelo Google Acadêmico.....	41
Figura 4 – Imagem da página inicial, do texto principal, da obra <i>Hippie</i> , de Paulo Coelho.....	47
Figura 5 – Visão geral dos arquivos das obras de Paulo Coelho em PDF, Word e TXT.....	49
Figura 6 – Visão geral dos arquivos dos livros de diversos autores em PDF, Word e TXT.....	51
Figura 7 – <i>Wordlist</i> , com <i>stoplist</i> , com as 20 primeiras palavras, de todas as obras de Paulo Coelho.....	53
Figura 8 – <i>Keyword</i> , com <i>stoplist</i> , com as 20 primeiras palavras-chave das obras dos romances literários de Paulo Coelho.....	54
Figura 9 – <i>Concord com as 20 primeiras ocorrências da palavra Deus</i> no conjunto de obras de Paulo Coelho.....	55
Figura 10 – <i>Concord com as 20 primeiras ocorrências da palavra demônio</i> no conjunto de obras de Paulo Coelho.....	56
Figura 11 – <i>Keyword</i> , com <i>stoplist</i> , com as 20 primeiras palavras-chave (substantivos e adjetivos) de todas as obras de Paulo Coelho, cujas chavidades constam em ordem crescentes na coluna <i>Keyness</i>	66

LISTA DE QUADROS

Quadro 1 – Obra <i>O Aleph</i> – Exclusões e correções das sílabas “fi”	47
Quadro 2 – Obra <i>O Aleph</i> – Supressões e correções das junções das letras “fi”	48
Quadro 3 – Obra <i>Hippie</i> – Alterações e correções do nome “Karla”	48
Quadro 4 – Obra “Hippie” – Conversões de aspas na sequência “ii”	48
Quadro 5 – Obra <i>Hippie</i> – Alterações da letra “m” para a junção das letras “rn”	48
Quadro 6 – Obra <i>O vencedor está só</i> – Correções das hifenizações.....	48
Quadro 7 – Montagem do <i>corpus</i> do Paulo Coelho – Atividades realizadas e tempo empregado.....	50
Quadro 8 – Montagem do <i>corpus</i> dos demais autores – Atividades realizadas e tempo empregado.....	52
Quadro 9 – Níveis de inteligibilidade de acordo com o índice Flesch.....	56

LISTA DE TABELAS

Tabela 1 – Tabela comparativa entre <i>O alquimista</i> , de Paulo Coelho, e obras de outros autores, em ordem alfabética pelos nomes dos autores.....	59
Tabela 2 – Tabela comparativa entre <i>O alquimista</i> , de Paulo Coelho, e as obras dos demais autores, apresentada em ordem crescente de índice Flesch.....	61
Tabela 3 – Tabela comparativa entre as obras de Paulo Coelho em ordem de alfabética de título.....	63
Tabela 4 – Tabela comparativa entre as obras de Paulo Coelho, apresentada em ordem crescente de índice Flesch.....	64

LISTA DE ABREVIATURAS E SIGLAS

ABL – Academia Brasileira de Letras

Capes – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

CT – Complexidade Textual

EC – Estilística de *Corpus*

EUA – Estados Unidos da América

FIT – Faculdades Integradas do Triângulo

Ileel – Instituto de Letras e Linguística

LC – Linguística de *Corpus*

Nilc – Núcleo Interinstitucional de Linguística Computacional (doravante Nilc)

Nilc-Metrix – Núcleo Interinstitucional de Linguística Computacional Metrix

PMU – Prefeitura Municipal de Uberlândia

OCR – *Optical Character Recognition*

PLN – Processamento de Linguagem Natural

STTR – *Standardised type/token ratio*

UFU – Universidade Federal de Uberlândia

USP – Universidade de São Paulo

Unitri – Centro Universitário do Triângulo

Volp – Vocabulário Ortográfico da Língua Portuguesa

WST – *WordSmith Tools*

SUMÁRIO

1 INTRODUÇÃO.....	17
2 REFERENCIAIS TEÓRICOS.....	21
2.1 LÉXICO	21
2.1.1 O léxico faz parte da história de cada comunidade	22
2.1.2 A constante modificação do léxico.....	24
2.1.3 Lexicologia.....	25
2.1.4 Algumas definições.....	27
2.1.4.1 Campos lexicais.....	27
2.1.4.2 Palavras gramaticais e lexicais.....	28
2.1.4.3 Acessibilidade textual.....	29
2.1.4.4 Complexidade textual.....	29
2.1.4.5 Densidade lexical.....	30
2.1.4.6 <i>Hapax legomena</i>	30
2.1.4.7 Palavras-chave.....	30
2.2 ESTILÍSTICA.....	30
2.2.1 Estilo.....	31
2.2.2 As teorias estilísticas de Charles Bally, Leo Spitzer e Amado Alonso.....	33
2.3 LINGUÍSTICA DE <i>CORPUS</i>.....	37
2.3.1 Um histórico da Linguística de <i>Corpus</i>	39
2.4 ESTILÍSTICA DE <i>CORPUS</i>.....	42
2.4.1 Estilística de <i>Corpus</i> e complexidade textual.....	44
3 METODOLOGIA DE PESQUISA.....	46
3.1 Preparação e montagem do <i>corpora</i>	46
3.1.1 <i>Corpus</i> das obras de Paulo Coelho.....	46
3.1.2 <i>Corpus</i> das obras dos demais autores	50
3.1.3 Usando o <i>WordSmith Tools</i>	52
3.1.4 Usando o índice Flesch.....	56
4 ANÁLISES E RESULTADOS.....	58
4.1 Comparativos entre <i>O alquimista</i> , de Paulo Coelho, e obras de demais autores.....	58
4.2 Comparativos entre as obras de Paulo Coelho.....	63

4.3 Análises das <i>hapax legomena</i> , palavras-chave e dos campos lexicais e semânticos das obras de Paulo Coelho.....	65
4.3.1 Campo lexical Religiosidades e Crenças.....	67
4.3.2 Campo lexical Geografia.....	68
4.3.3 Campo lexical Sentimentos.....	68
4.3.4 Campo lexical Anatomia.....	69
4.3.5 Campo lexical Astronomia.....	69
4.4 Análises das palavras <i>Deus e demônio</i> e seus cotextos no conjunto das obras de Paulo Coelho.....	69
4.4.1 <i>Deus</i> e seus cotextos no conjunto das obras de Paulo Coelho.....	70
4.4.2 <i>Demônio</i> e seus cotextos no conjunto das obras de Paulo Coelho.....	70
4.4.3 <i>Deus e demônio</i> no estilo lexical de Paulo Coelho.....	71
5 CONSIDERAÇÕES FINAIS.....	72
REFERÊNCIAS.....	75
APÊNDICES.....	81
Apêndice A – Estatísticas apresentadas pelo <i>WordSmith Tools</i> para as obras de diversos autores.....	81
Apêndice B – Estatísticas apresentadas pelo <i>WordSmith Tools</i> para as obras de Paulo Coelho.....	83
Apêndice C – Índices Flesch apresentados para as 17 obras de autores diversos.....	85
Apêndice D – Índices Flesch apresentados para as obras de Paulo Coelho.....	88

1 INTRODUÇÃO

A proposta metodológica para análises linguísticas de obras literárias apresentada nesta dissertação justifica-se ao promover uma conexão entre a Linguística e a Literatura. A ideia surgiu por acreditarmos que tal união só acrescenta benefícios para ambas as áreas, indissociáveis de acordo com nosso entendimento. Desta forma, pretendíamos realizar pesquisas envolvendo análises linguísticas de diversos autores e suas obras literárias, utilizando-nos de uma metodologia possível graças à Linguística. No desenvolvimento da proposta usamos teorias das áreas do Léxico, a Estilística e a Estilística de *Corpus* e como metodologia a Linguística de *Corpus*.

O léxico se revela importante nas análises literárias pois é através dele que os escritores contam suas histórias, utilizando um palavrado de acordo com suas intenções e criações. Quanto à Estilística, pelos seus pressupostos temos consciência do estilo utilizado pelo autor para obter efeitos estéticos e psicológicos junto aos leitores. Tal estilo também revela as características textuais do autor, suas escolhas linguísticas e, em algumas ocasiões, sua história de vida e aspectos psicológicos. Por sua vez, a Linguística de *Corpus* é uma eficaz e ágil metodologia de análise. Nos valem ainda da Estilística de *Corpus*, que permite realizar análises unindo Estilística e Linguística de *Corpus*.

Como principal objeto de estudo para testar a metodologia, escolhemos os romances literários de Paulo Coelho, sendo eles, em ordem de lançamento: *O diário de um mago* (1987), *O alquimista* (1988), *Brida* (1990), *As Valkírias* (1992), *Na margem do rio Piedra eu sentei e chorei* (1994), *O Monte Cinco* (1996), *Verônica decide morrer* (1998), *O demônio e a srta. Prym* (2000), *Onze minutos* (2003), *O Zahir* (2005), *A bruxa de Portobello* (2006), *O vencedor está só* (2008), *O Aleph* (2010), *Adultério* (2014), *A espiã* (2016) e *Hippie* (2018).

Também nos valem de 17 obras famosas da literatura brasileira, escritas por outros autores, para contrastar em alguns aspectos com a obra *O alquimista*, de Paulo Coelho, sendo elas, também em ordem de lançamento: *A carta*, de Pero Vaz de Caminha (1500); *Auto representado na Festa de São Lourenço*, de Padre José de Anchieta (1583); *Sermão de Santo Antônio aos peixes*, de Padre Antônio Vieira (1682); *Seleção de obras poéticas*, de Gregório de Matos (antologia, diversos anos); *Marília de Dirceu*, de Tomás Antônio Gonzaga (1792); *Iracema*, de José de Alencar (1865); *Lira dos vinte anos*, de Álvares de Azevedo (1853); *Espumas flutuantes*, de Castro Alves (1870); *Dom Casmurro*, de Machado de Assis (1899); *O cortiço*, de Aluísio Azevedo (1890); *Alma inquieta*, de Olavo Bilac (antologia, diversos anos);

Broquéis, de Cruz e Sousa (1893); *Os sertões*, de Euclides da Cunha (1902), *Macunaíma: o herói nenhum caráter*, de Mário de Andrade (1928), *Vidas secas*, de Graciliano Ramos (1938), *Grande sertão: veredas*, de Guimarães Rosa (1956) e *Budapeste*, de Chico Buarque (2003).

Quanto a Paulo Coelho, o autor é um caso atípico, uma vez que mesmo sendo um dos autores mais famosos da atualidade, ainda não tem grande apreço por parte dos críticos literários e dos acadêmicos. Como menciona o escritor Maicon Tenfen, em sua coluna na revista *Veja*, publicada em 24 de agosto de 2017 e intitulada *Aos 70 anos, Paulo Coelho ainda é o nosso maior péssimo escritor*¹:

É claro que a crítica sempre detestou o mago. Parágrafos mal escritos, pobreza vocabular, diálogos artificiais, personagens estereotipadas, todos os defeitos imagináveis foram atribuídos aos seus livros. A academia, então, é melhor nem mencionar. Mestrandos e doutorandos que se atreveram a estudar a obra de Paulo Coelho receberam vaias em congressos e simpósios sobre literatura.

No entanto, é quase impossível nos referirmos à cultura e à literatura contemporâneas brasileiras e mundiais sem mencionarmos Paulo Coelho. Segundo a Academia Brasileira de Letras (ABL)², o autor iniciou suas produções literárias em 1982, quando escreveu *Arquivos do inferno*, que não obteve sucesso. Em 1985 colaborou com a escrita de *O manual prático do vampirismo*, que recolheu por considerá-lo de má qualidade. O sucesso chegou em 1987 com o lançamento de *O diário de um mago*. Desde então, seus livros já venderam mais de 250 milhões de exemplares em 170 países³ e foram traduzidos para 81 idiomas⁴. Paulo Coelho ainda assinou contratos com emissoras e produtoras nacionais e internacionais de TV e de cinema, interessadas em adaptar seus livros em obras televisivas ou cinematográficas, foi empossado membro da ABL em junho de 2002 e recebeu 33 prêmios e homenagens entre os anos de 1995 e 2008⁵. Como a página da ABL menciona dados sobre o autor somente até 2008, realizamos diversas pesquisas para citarmos também os prêmios dos últimos anos, mas os endereços eletrônicos visitados também utilizam a página da ABL como referência,

¹ Disponível em: <https://veja.abril.com.br/blog/o-leitor/aos-70-anos-paulo-coelho-ainda-o-nosso-maior-pessimo-escriptor/>. Acesso em 20 maio 2020.

² Disponível em: <https://www.academia.org.br/academicos/paulo-coelho>. Acesso em 1 jun 2020.

³ Disponível em: <http://g1.globo.com/fantastico/noticia/2018/05/hippie-traz-paulo-coelho-em-aventura-num-onibus-de-amsterda-para-o-nepal.html>. Acesso em 17 set 2021.

⁴ Disponível em: <https://cultura.estadao.com.br/noticias/literatura,em-hippie-paulo-coelho-revive-historias-de-amor-e-de-tortura-fisica,70002314781>. Acesso em 17 set 2021.

⁵ Disponível em: <https://www.academia.org.br/academicos/paulo-coelho>. Acesso em 17 set 2021.

também se limitando até 2008. O autor também foi mencionado no *Guinness Book of Records*, atual *Guinness World Records*, como o escritor que autografou, em uma só sessão, o maior número de traduções (53 no total), de uma única obra, neste caso, *O alquimista*, durante a Feira Internacional do Livro, em Frankfurt, na Alemanha, em 2003⁶.

O maior sucesso editorial de Paulo Coelho, *O alquimista*, ficou em primeiro lugar na lista de mais vendidos em 18 países⁷, permaneceu durante sete anos (364 semanas) na lista de mais vendidos do jornal *The New York Times*⁸ e foi apontado pelo *Guinness World Records* como o livro mais traduzido e vendido de um autor vivo no mundo⁹. De acordo com uma pesquisa divulgada pelo jornal *O Globo*, em 2021, *O alquimista* é o livro mais traduzido na história da literatura brasileira e nono na literatura mundial¹⁰.

Segundo a ABL, Paulo Coelho também atuou como jornalista e compositor, além de diretor, professor e autor de teatro. No cenário musical, foi autor de mais de diversas composições, muitas em parceria com o músico Raul Seixas, como *Gita*, *Al Capone* e *Eu nasci há dez mil anos atrás*¹¹.

Mesmo com as controvérsias entre sucesso e falta de reconhecimento, Paulo Coelho atrai inúmeros leitores, principalmente de seus romances literários. Diante disso, é de se imaginar que ele conte com recursos capazes de despertar o interesse das pessoas. Por se tratar de obras literárias, o meio utilizado pode ser, mesmo que inconscientemente pelo autor, seu estilo lexical, com características capazes de explicar tão grande sucesso.

Diante disso, utilizamos os romances literários de Paulo Coelho como o principal objeto de estudo da proposta de análises desta pesquisa por serem as principais produções que o tornaram um sucesso mundial. Nos valemos ainda de outro *corpus* composto por obras famosas da literatura brasileira, promovendo um contraste entre elas e *O alquimista*, de Paulo Coelho, quanto a itens como densidade lexical e complexidade textual.

Para o desenvolvimento de nossa pesquisa, partimos do seguinte questionamento:

⁶ Disponível em: <http://www.academia.org.br/academicos/paulo-coelho/biografia>. Acesso em 1 jun 2020.

⁷ Disponível em: <http://www.academia.org.br/academicos/paulo-coelho/biografia>. Acesso em 17 set 2021.

⁸ Disponível em: <https://entretenimento.uol.com.br/noticias/redacao/2015/08/21/paulo-coelho-comemora-recorde-de-7-anos-na-lista-de-mais-vendidos-do-nyt.htm>. Acesso em 17 set 2021.

⁹ Disponível em: <https://www1.folha.uol.com.br/ilustrada/2018/04/o-politicamente-correto-vai-matar-a-literatura-diz-paulo-coelho.shtml>. Acesso em 17 set 2021.

¹⁰ Disponível em: <https://blogs.oglobo.globo.com/afonso-borges/post/o-alquimista-de-paulo-coelho-e-o-livro-mais-traduzido-da-historia-do-brasil.html>. Acesso em 27 dez 2021.

¹¹ Disponível em: <https://www.academia.org.br/academicos/paulo-coelho/biografia>. Acesso em 17 set 2021.

- Quais são as ferramentas computacionais possíveis para fazer análises linguísticas de obras literárias?

Diante disso, apresentamos uma proposta, tendo como metodologia a Linguística de *Corpus* e promovendo uma conexão entre a Linguística e a Literatura.

Quanto ao nosso objetivo geral nesta pesquisa, este foi propor uma metodologia de análises linguísticas de obras literárias, capaz de verificar aspectos lexicais e estilísticos de um ou mais autores e de suas obras, tendo como pressupostos teóricos estudos sobre o Léxico, a Estilística, a Linguística de *Corpus* e a Estilística de *Corpus* e como metodologia a Linguística de *Corpus*.

Como mencionamos anteriormente, as análises foram feitas utilizando um *corpus* com os 16 romances literários de Paulo Coelho e outro *corpus* com 17 obras famosas da literatura brasileira para observar se o estilo lexical de Paulo Coelho possui marcas próprias que contribuem para o seu sucesso.

A pesquisa conta ainda com objetivos específicos, sendo eles:

- Elaborar uma proposta de metodologia de análises linguísticas de obras literárias;
- Verificar se o estilo lexical de Paulo Coelho permite uma melhor acessibilidade textual às suas obras, verificando itens como palavras-chave dos romances literários do autor por meio de um recorte com substantivos e adjetivos, campos lexicais do conjunto de suas obras e uso de *hapax legomena*, além da complexidade textual e da densidade lexical de suas obras;

- Estabelecer comparativos entre a densidade lexical e a complexidade textual de *O alquimista*, de Paulo Coelho, e obras de outros autores, lançadas em diferentes épocas e escolas literárias.

Finalmente, destacamos que esta dissertação é apresentada em cinco capítulos. O primeiro é este, no qual apresentamos as questões pertinentes à pesquisa, sendo elas, a introdução, as justificativas, a pergunta de pesquisa, o objetivo geral e os objetivos específicos. No segundo fizemos a exposição dos referenciais teóricos que amparam a pesquisa, por sua vez, sobre Léxico, Lexicologia, Estilística e Estilística de *Corpus*; e ainda a teoria da metodologia utilizada, ou seja, a Linguística de *Corpus*. No terceiro descrevemos os passos metodológicos adotados com o uso da ferramenta *WordSmith Tools* e do índice Flesch. No quarto expomos as análises e os resultados obtidos juntos aos *corpora* com os romances de Paulo Coelho e as obras dos demais autores. No quinto constam nossas considerações finais.

2 REFERENCIAIS TEÓRICOS

Neste capítulo apresentamos os referenciais teóricos nos quais nos respaldamos para elaboração da metodologia, abordando estudos sobre Léxico, Estilística, Linguística de *Corpus* e Estilística de *Corpus*.

2.1 LÉXICO

O léxico de uma língua natural é fundamental ao homem, pois permite a eficácia na transmissão e recepção do conhecimento. Quando alguém conhece algo novo, precisa assimilar esta nova realidade, seja de forma autônoma ou transmitida por alguém ou algum meio de informação. Esta novidade precisa ser rotulada e utilizada por uma comunidade linguística, pois é inimaginável uma comunicação ocorrer sem uma nomeação unificada. É diante desta necessidade que surge o léxico de uma língua natural, ou seja, um conjunto comum de palavras usadas por um grupo de pessoas para se comunicarem de forma inteligível. Esta concepção do léxico é baseada, principalmente, a partir de uma definição de Biderman que afirma ser o léxico “a etapa primeira no percurso científico do espírito humano de conhecimento do universo” (BIDERMAN, 2001, p. 13). Neste processo de registrar, codificar, interiorizar e classificar conhecimentos a partir de realidades vividas, percebe-se ainda como as pessoas entendem o mundo onde habitam.

Ao reunir os objetos em grupos, identificando semelhanças e, inversamente, discriminando os traços distintivos que individualizam esses seres e objetos em entidades diferentes, o homem foi estruturando o mundo que o cerca, rotulando essas entidades discriminadas. Foi esse processo de nomeação que gerou o léxico das línguas naturais. (BIDERMAN, 2001, p. 13)

Ao criar nomenclaturas espontaneamente, o homem confere individualidades a tudo que existe. Em dois textos nos quais aborda tal assunto, Biderman (1987 e 2001) considera esta etapa como uma estratégia engenhosa de simbolização, surgida durante o dia a dia e influenciada pelas tradições, costumes, acontecimentos, enfim, tudo o que faz parte da história e da rotina das pessoas.

Em suma, o universo conceptual de uma língua natural pode ser descrito como um sistema ordenado e estruturado de categorias léxico-gramaticais. As palavras geradas por tal sistema nada mais são

do que rótulos, através dos quais o homem interage cognitivamente com o seu meio. Vale a pena insistir no fato de que as categorias lexicais variam de língua para língua, raramente ocorrendo que dois idiomas sejam dotados dos mesmos tipos categoriais. (BIDERMAN, 2001, p. 14)

Além de ser mutante, o léxico tem um heterogêneo e democrático processo de formação. Segundo Vilela (1997), a sua constituição não é feita de forma hermética, permitida apenas para determinadas categorias, como linguistas e escritores. Pelo contrário, ocorre uma união inconsciente entre especialistas e pessoas de diversas classes sociais, profissões, religiões, etc. O léxico ainda abriga ambiguidades, como arcaísmos convivendo com neologismo, linguagem formal usada em conjunto com coloquialismos, enfim, uma infinidade de combinações que confere múltiplas características a ele.

2.1.1 O léxico faz parte da história de cada comunidade

Ao longo de sua história, o homem desenvolveu formas cada vez mais eficazes de se comunicar e o léxico sempre teve função primordial neste processo. Segundo Biderman (2001), ao revelar o seu meio de origem, o léxico de uma língua natural permite um contato com diversas características e patrimônios de uma comunidade linguística, como vocabular, cultural e histórico.

Abbate (2011) corrobora Biderman (2001) ao mencionar que ao entrarmos em contato com o léxico de uma língua natural, temos acesso a um retrato cultural da sociedade que o utiliza. O léxico torna perceptível aspectos como cultura, tradição, religiosidades, crenças, costumes, escolaridades, ou seja, uma infinidade de peculiaridades. Mesmo que cada pessoa faça uso do léxico de uma forma individual, isto não será totalmente independente, pois, paralelamente, também será usado um palavreado coletivo, a fim de ser compreendido pelas demais pessoas.

Cada palavra selecionada nesse processo acusa as características sociais, econômicas, etárias, culturais... de quem a profere. Partindo dessa premissa, estudar o léxico de uma língua é abrir possibilidades de conhecer a história social do povo que a utiliza. (ABBADE, 2011, p. 1332).

Tomamos a liberdade de imaginar como seria a transferência do conhecimento em uma sociedade onde cada pessoa possuísse um léxico individual. Entendemos que isto seria

impossível. Seria o mesmo que conversar com alguma pessoa que se utiliza de um idioma com o qual nunca tivemos contato. Tal fala vira um emaranhado indefinido, uma espécie moderna da Torre de Babel¹², e não conseguimos, sequer, saber quando começa ou termina uma palavra.

Outro aspecto importante do léxico diz respeito à sua existência, que ocorre graças ao *thesaurus*, ou seja, “uma herança de signos lexicais herdados e de uma série de modelos categoriais para gerar novas palavras” (BIDERMAN, 1987, p. 83). Isto evidencia outra característica, a enorme dificuldade de se criar palavras que não tenham vestígios de outras já existentes. Assim sendo, o *thesaurus* é indispensável para o dinamismo lexical, pois mesmo que os indivíduos criem novas palavras, estas serão provenientes de um modelo formal preexistente.

Além da herança lexical, segundo J. Rey Debove (*apud* Biderman, 1987), normalmente uma pessoa memoriza aproximadamente 20 mil palavras e seu uso é extremamente complexo, porém precisa ser ágil e eficaz. Diante disso, segundo Biderman (1987), a memória de cada pessoa se torna um importante mecanismo para fazer com que a comunicação seja compreendida. Ao entrar rotineiramente em contato com várias palavras, as pessoas memorizam algumas, criando uma espécie de arquivo lexical individual, e as utilizam de acordo com as ocasiões.

É evidente que nenhum indivíduo possui um banco de dados tão volumoso na sua memória. (...) Se somarmos a esse obstáculo numérico o problema que resulta da complexidade combinatória desses elementos, é necessário supor que o cérebro humano precisa dispor de um sistema operacional extremamente eficiente na manipulação desses dados. Suponho que o cérebro organiza uma estruturação dos lexemas de grande funcionalidade para que, em milésimos de segundo, possa recuperar não só o significado de uma palavra, mas também todas as suas características gramaticais e os usos que lhe são adequados conforme o contexto do discurso, o tipo de discurso, a situação momentânea e o registro lingüístico requerido pela situação, pelo interlocutor e pelo assunto. Daí a necessidade absoluta de ordenação lógica do léxico; caso contrário, o indivíduo jamais poderia recuperar na sua memória um lexema específico para uma sentença particular que quer construir, ou que ouviu ou leu e deve decodificar. Em função dessa necessidade e da configuração do

¹² Mencionamos a Torre de Babel (*Bíblia*, 1990), devido à narração bíblica constante no *Antigo Testamento*, no livro *Gênesis* (11,1-9), que conta a história da tentativa de construção de uma torre, pelos descendentes de Noé, de altura suficiente para alcançar o Céu. A ato teria sido considerado uma soberba por Deus que castigou os homens confundindo suas línguas e fazendo com que seus dizeres se tornassem incompreensíveis entre tal grupo de pessoas.

patrimônio lingüístico herdado, o léxico se estrutura segundo padrões hierárquicos e sistemáticos. (BIDERMAN, 1987, p. 83)

Assim, entendemos que, fazendo uma analogia a um labirinto com seus diversos caminhos, o indivíduo que deseja se comunicar se vê diante de várias opções lexicais que fazem parte de sua memória. De forma inconsciente e instantânea, ele utiliza palavras que melhor lhe convém para se expressar.

2.1.2 A constante modificação do léxico

O léxico é modificado constantemente devido às influências, principalmente, sociais e culturais. Biderman (2001) menciona a acelerada forma como ocorre a renovação lexical e que, de acordo com J. Rey-Debove, a renovação de um vocabulário é de aproximadamente 10% no prazo de 25 anos, sendo que um idioma pode atingir e ultrapassar a marca de 400 mil palavras. Este dinamismo do léxico ocorre devido a vários motivos, como expansão, contração, marginalização, desaparecimento e, até mesmo, ressurgimento de palavras.

Abbade (2011) também destaca o dinamismo do léxico e a importância de realizar estudos da língua considerando a sua evolução. Segundo a autora, quando isto não ocorre, pode acontecer de o estudo se mostrar ineficiente.

Outro autor que cita o dinamismo lexical é Pereira (1932), *apud* Barbosa e Marine (2011), que menciona ser fundamental a mobilidade do léxico. Ao fazer uma analogia com a vida humana, assim, como Biderman (2001) e Abbade (2011), Pereira menciona que, da mesma forma que o homem, as palavras nascem, tem suas vidas, passam por transformações, morrem e algumas até ressuscitam.

Neste cenário, as influências dos usuários da língua também se revelam capazes de alterar o léxico. Durante o uso das palavras, os indivíduos atribuem conotações particulares a elas, geralmente características dos seus meios sociais, econômicos e culturais. Desta forma, eles agem sobre a estrutura lexical, muitas vezes alterando as suas significações. Biderman (2001) afirma que os indivíduos são capazes de gerar e modificar os significados e as interpretações das palavras, principalmente aqueles mais criativos e de maior competência lingüística, como os escritores e os poetas.

A constante mudança lexical deve-se à percepção de novas realidades e ao desenvolvimento das ciências, entre outros motivos. Isto faz com que comunidades em consideráveis estágios de civilização ampliem constantemente seu léxico, criando rótulos e

obrigando as pessoas a se adaptarem às novas realidades. Biderman (2001) cita como exemplo o léxico das línguas modernas, constantemente alterado devido ao contínuo progresso técnico e científico, à progressiva integração das culturas e dos povos e à rapidez das mudanças sociais causadas, principalmente, pela comunicação de massa.

Outro aspecto importante nas mudanças lexicais são os empréstimos linguísticos. Biderman (1987) evidencia o impacto deles, a maioria proveniente do idioma inglês, utilizados geralmente por cientistas e técnicos em suas criações neológicas. Embora dominante pelo fato de ser considerado o idioma universal em áreas como Economia, Ciência e Tecnologia, além da hegemonia mundial dos Estados Unidos da América (EUA) que o utiliza, Biderman (2001) destaca que o inglês não é o único idioma exportado. Diversos outros, como o francês e o latim, tiveram palavras incorporadas fora de suas regiões de origem. A autora afirma que no caso da Língua Portuguesa houve a incorporação de diversas palavras de outras línguas, sendo que muitas foram adaptadas ou, até mesmo, recriadas.

Apresentamos um fato sobre dinamismo lexical ocorrido no Brasil: a última atualização do Vocabulário Ortográfico de Língua Portuguesa (Volp), lançada pela ABL em julho de 2021¹³. O material recebeu o acréscimo de 1.160 palavras, como criptomoeda, feminicídio, telemedicina, ciberataque, botox e bullying. Algo observado foi a introdução de 65 palavras (5,6%) ligadas direta ou indiretamente à pandemia de Covid-19, como lockdown e home office¹⁴.

Após estas abordagens gerais sobre léxico, nosso próximo tópico aborda a Lexicologia, importante para o desenvolvimento da nossa metodologia de análises linguísticas de obras literárias.

2.1.3 Lexicologia

Os estudos envolvendo as palavras datam da Antiguidade Clássica, entretanto, a Lexicologia, que consiste em uma ciência que “realiza o estudo científico do léxico” (Genouvrier *apud* Barbosa, 1991, p. 185) é considerada recente.

Em seu percurso, inicialmente a Lexicologia foi desprezada. “Quase nada se fazia com as palavras de uma língua além de organizá-las alfabeticamente e buscar suas definições a

¹³ Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2021-07/academia-brasileira-de-letras-lanca-nova-edicao-online-do-volp>. Acesso em 1º nov 2021.

¹⁴ Disponível em: <https://educacao.uol.com.br/noticias/2021/10/17/pandemia-coronavirus-novas-palavras-dicionario.htm>. Acesso em 1º nov 2021.

partir de sua literatura.” (ABBADE, 2011, p. 1.333). Segundo a autora, a mudança ocorreu no final do século XIX e a Linguística deixou de ser direcionada principalmente à investigação fonética, passando a ser mais utilizada no campo das pesquisas lexicais. Desde então, a Lexicologia foi expandida, oferecendo ampla quantidade de conhecimentos para outras ciências, como a Glotocronologia (método de investigação que calcula a data de separação de línguas aparentadas), a Semântica (estudos sobre a significação das palavras), a Etnolinguística (estudos sobre a relação entre a linguagem e aspectos socioculturais), a Psicolinguística (estudos sobre o processo de aquisição de uma língua nos aspectos linguístico e psicológico) e a Neurolinguística (pesquisas sobre as relações entre a estrutura cognitiva do ser humano e sua capacidade linguística). Ressaltamos que estas breves definições apresentadas tiveram como fonte o *Michaelis Moderno Dicionário da Língua Portuguesa*¹⁵.

Atualmente a Lexicologia tem sua valorização reconhecida no campo das ciências da linguagem, sendo definida como o “o ramo da Linguística que se dedica ao estudo científico do léxico, uma ciência que estuda tanto as unidades lexicais como a organização do léxico” (BEILKE, 2016, p. 62).

Quanto às funções da Lexicologia, Barbosa (1991) e Abbade (2011) são unânimes em destacar que a ciência compreende diversos domínios, entre eles formação, criação e importação de palavras e análises das relações lexicais de uma língua com os meios natural, social e cultural aos quais ela pertence e é utilizada. A ciência também é capaz de revelar características ideológicas, culturais e sociais dos usuários da língua, além de possibilitar estudos das criações e renovações lexicais, como os neologismos. Ressaltamos que a lista de possibilidades de trabalhos com a Lexicologia é ampla e sempre está em processos de renovação e ampliação.

Os pressupostos teóricos da Lexicologia permitiram diversas práticas, garantindo ganhos positivos para a sociedade. Barbosa (1991) menciona que a Lexicologia Descritiva possibilitou o surgimento da Lexicologia Aplicada, que consiste na utilização da ciência em diversos campos, tecnologias e práticas, como o ensino de língua materna ou estrangeira, os procedimentos para realizar traduções automáticas ou mecânicas, as técnicas de documentação e ainda as análises e tratamentos dos distúrbios da linguagem. A Lexicologia também contribuiu com o surgimento de outras ciências, como a Estilística, que abordaremos posteriormente nesta pesquisa.

¹⁵ Disponível em: <https://michaelis.uol.com.br/moderno-portugues/>. Acesso em 17 out 2021.

2.1.4 Algumas definições

Antes de encerrarmos as abordagens sobre léxico, explicamos algumas expressões utilizadas nesta pesquisa.

2.1.4.1 Campos lexicais

A teoria dos campos lexicais define que as palavras são reunidas e organizadas conforme suas dependências e semelhanças. De acordo com Abbade (2011) para melhor entender o significado de cada palavra de um enunciado ou em determinada posição do enunciado, é preciso considerar o todo no qual ela está expressa.

Os campos lexicais representam uma estrutura, um todo articulado, onde há uma relação de coordenação e hierarquia articulada entre as palavras que são organizadas à maneira de um mosaico: o campo léxico. As palavras são organizadas em um campo com mútua dependência, adquirindo uma determinação conceitual a partir da estrutura do todo. (ABBADE, 2011, p. 1332)

Conforme Biderman (2001), o léxico se assemelha a um labirinto, formado por diversas significações linguísticas, tornando quase impossível definir o significado de uma palavra sem considerar as demais que compõem o enunciado.

Em Abbade (2011) tomamos conhecimento da importância da oposição entre as palavras, sendo que o significado de uma depende do significado da outra. Assim como em uma corrente composta por gomos, o léxico se constitui através de elos que, se alterados, causam mudanças no significado do enunciado.

O significado de cada palavra vai depender do significado de suas vizinhas conceituais. Elas só têm sentido como parte de um todo, pois só no campo terão significação. Assim, para entender a lexia individualmente é necessário observá-la no seu conjunto de campo, pois fora desse conjunto não pode existir uma significação, uma vez que a mesma só existe nesse conjunto e em sua razão. (ABBADE, 2011, p. 1332)

Em seu *Curso de Linguística Geral*, Saussure (2012) também aborda a importância das palavras serem inter-relacionadas, obrigando-as a assumirem seus valores. De acordo com

Saussure, o mecanismo linguístico é baseado em identidades e diferenças, com termos e valores solidários.

Assim, o valor de qualquer termo que seja está determinado por aquilo que o rodeia; nem sequer da palavra que significa “sol” se pode fixar imediatamente o valor sem levar em conta o que lhe existe em redor; línguas há em que é impossível dizer "sentar-se ao sol". (SAUSSURE, 2012, p. 163)

Saussure afirma que os valores são possíveis graças aos diferenciais de cada palavra. Eles não são definidos pelas suas semelhanças, mas por suas dessemelhanças com os demais, ou seja, por suas relações negativas com outras palavras do sistema. Sendo assim, é certo dizer que uma forte característica da palavra é ser o que a outra não é.

Abbade (2011) menciona que Coseriu também propôs uma análise do vocabulário amparada nas oposições entre as palavras, porém, observando a funcionalidade da língua em sua utilização. Para Coseriu, “o importante é apresentar o funcionamento real da língua.” (*apud* ABBADE, 2011, p. 1342).

Finalmente, encerramos esta abordagem sobre campos lexicais corroborando Abbade (2011), que afirma que a teoria dos campos lexicais ainda encontra-se em um certo grau de dificuldade quanto a apresentar concisão sobre a estruturação dos campos lexicais.

2.1.4.2 Palavras gramaticais e lexicais

Iniciamos este item definindo o que é palavra. Conforme Abbade (2011, p. 1333), “é um termo genérico, tradicionalmente utilizado na língua, fazendo parte do vocabulário de todos os falantes.”

Segundo Martins (2012), as palavras se dividem entre gramaticais e lexicais. Quanto às gramaticais, suas significações dependem dos contextos linguísticos nos quais são utilizadas. Embora sejam importantes e apareçam frequentemente, elas são pouco numerosas. “Sua função pode estar relacionada com o ato de enunciação, com a organização do discurso ou texto ou com a estruturação da frase” (MARTINS, 2012, p. 99). Como alguns exemplos de palavras gramaticais, citamos os pronomes, os artigos, os numerais e as conjunções.

Quanto às palavras lexicais, Martins (2012) explica que elas têm significação por si mesmas e despertam representações mentais, sendo também chamadas como lexias, lexemáticas, autossemânticas, lexicográficas, reais ou plenas. “Diz-se que elas têm significação

extralinguística ou externa, visto que remetem a algo que está fora da língua e que faz parte do mundo físico, psíquico ou social.” (MARTINS, 2012, p. 104). As palavras lexicais aparecem em número indeterminável e sempre se renovam, causando constantes alterações da língua onde estão inseridas. A lista delas abrange os substantivos, os adjetivos, os advérbios e os verbos que exprimem ação ou processo mental.

Após esta explanação, abordaremos a seguir a Estilística, área que nos permitiu analisar o estilo lexical de Paulo Coelho, presente em seus romances literários.

2.1.4.3 Acessibilidade textual

A acessibilidade textual é a capacidade de entendimento do conteúdo de um texto por parte de uma pessoa. Segundo Finatto (2017), as pesquisas sobre tal tema tem como principal motivo entenderem como um texto consegue ou não atingir seus objetivos junto ao público ao qual é destinado.

2.1.4.4 Complexidade textual

A definição de complexidade textual pode ser dita como propriedade ou condição de um texto diante da familiaridade do leitor com o material, podendo ser observada diante de sua construção verbal e configuração gramatical. Conforme Paraguassu (2018), a percepção quanto à complexidade textual de cada texto depende do público que o recebe. Isto porque, para determinadas pessoas, um enunciado pode ser de difícil ou, até mesmo, impossível compreensão, enquanto que para outras pode ser de fácil inteligibilidade. Para cada leitor existem áreas de maior ou menor contato, por exemplo, para uma pessoa da área de Medicina, um laudo de exame é facilmente compreendido, enquanto que uma pessoa da área de Direito, talvez, tenha dificuldades de compreender o laudo.

Finatto (2011) menciona uma série de aspectos que podem ser observados para se analisar a complexidade de um texto, como léxico utilizado, extensão do texto e número de palavras por frases, entre outros. A autora destaca que estas análises podem ser feitas por recursos oferecidos pela informática, entretanto, os resultados apresentados são estimativas e não devem ser tratados como inquestionáveis, sendo necessário realizar testes diretos com os leitores.

2.1.4.5 Densidade lexical

A densidade lexical, segundo Toledo (2017), é a proporção de itens lexicais utilizada pelo falante ou escritor, sendo “mensurada equacionando-se a relação entre o total de itens lexicais e o número total de palavras no texto produzido pelo indivíduo. (TOLEDO, 2017, p. 37). Diante disso, quando mais alta a densidade, mais variado o léxico utilizado.

2.1.4.6 *Hapax legomena*

Nos valem da definição apresentada por Fromm, Santos, Grama e Beilke (2020) de que são palavras que aparecem somente uma vez no *corpus*.

2.1.4.7 Palavras-chave

Conforme Sardinha (2009), as palavras-chave são obtidas por meio de análises estatísticas, nas quais se verifica o total de vezes que cada palavra aparece no *corpus* de pesquisa e no *corpus* de referência. Observa-se quais palavras são usadas em maior ou menor frequência quando comparadas ao número total de palavras do texto.

2.2 ESTILÍSTICA

A Estilística é a ciência da linguagem que estuda os estilos das comunicações, principalmente dos textos, além dos efeitos estéticos e expressivos obtidos junto aos destinatários, esclarecem Morejón e Martins (2001).

Ao produzir um texto, o autor tem à sua disposição diversas formas de expressão e as utiliza de acordo com sua personalidade, público-alvo e influências linguísticas, psicológicas e sociais, entre outros motivos e escolhas que consistirão em seu estilo. O estilo, portanto, “consiste na linguagem pessoal, idiossincrática, por oposição à linguagem de todos”. (DUARTE, 2006, p. 45)

Conforme Morejón e Martins (2001), a qualidade resultante de tais escolhas se deve à necessidade de tornar o enunciado o mais eficaz e interessante possível. “A Estilística como dissemos, procura chegar a decifrar os enigmas da obra poética, através da investigação do que comumente se vem denominando estilo.” (MOREJÓN; MARTINS, 2001, p. 156).

Nesta investigação, que consiste nos estudos estilísticos, o pesquisador não precisa desprezar a crítica e a história literária e se fixar apenas no material que tem à sua disposição. Morejón e Martins (2001) ressaltam que pode-se fazer o contrário, usando como referências os valores que interferiram na produção da obra, como a história de vida do autor, os locais onde ele viveu, os fatos que constaram no seu dia a dia, suas referências que lhe serviram de inspiração e, até mesmo, seus aspectos psicológicos, entre outros.

O investigador do estilo deve ser uma alma aberta a todos os horizontes, um espírito vivo e alerta ante o mundo de sensações, emoções e volições que a obra literária possa despertar, e quanto mais rica for sua capacidade de compreensão e sua cultura, sua intuição e suas possibilidades expressivas, tanto maior será a sua facilidade para descobrir os mistérios da arte. (MOREJÓN; MARTINS, 2001, p. 157, 158).

Percebe-se claramente que um dos principais objetivos da Estilística é analisar o estilo do enunciador, ou seja, as escolhas feitas por ele dentre os elementos linguísticos disponíveis. Processo que resulta em efeitos estéticos e expressivos com traços de sua individualidade.

2.2.1 Estilo

A classificação da Estilística como uma ciência da linguagem que se ocupa dos estudos sobre o estilo nos remete às definições de estilo. Segundo Martins (2012), a origem da palavra estilo advém dos tempos antigos, quando se usava o *stilus*, um instrumento pontiagudo para escrever em tábuas, originando o que atualmente é designado como estilo.

Diversos pesquisadores consideram estilo apenas quando se trata de textos analisados por um viés literário. Outros defendem que ele deve ser observado nos diversos usos da língua e relacionado aos autores, às obras e/ou às reações dos leitores diante das obras. Corroboramos a definição de Viegas (1982, p. 178) que afirma que “como todo esforço de criação, estilo pode ser igualmente entendido como resultado de um trabalho”.

Pela definição de Duarte (2006), estilo é um artifício pelo qual se criam condições que contribuem para se atingir o efeito almejado. Devido a isto, estilo não deve ser analisado considerando apenas a beleza, mas todos os motivos e artifícios que impactam o público.

Estes artifícios aos quais Duarte se refere são chamados por Morejón e Martins (2001) como escolha quanto aos procedimentos disponibilizados pelo sistema linguístico e usados pelos falantes.

Esta escolha, e a qualidade dela resultante, são ditadas pela necessidade de tornar o enunciado tão eficaz quanto seja possível. Utilizar-se-ão as possibilidades expressivas, afetivas, evocadoras e estéticas da língua para obter o matiz e os efeitos desejados. (MOREJÓN; MARTINS, 2001, p. 156)

Ao se referir a estilo, Fiorin (2000) o considera como “o conjunto de traços particulares que define desde as coisas mais banais até as mais altas criações artísticas.” (FIORIN, 2000, p. 40). Fiorin destaca que estilo pode ser considerado como características que expressam a particularidade da obra, do autor e da época de sua produção, entre outras. Sendo pela oposição ao outro e pela sua heterogeneidade, que o estilo se constitui.

O termo *estilo* alude, então, a um fato diferencial: diferença de um autor em relação a outro, de um pintor relativamente a outro, de uma época em relação a outra, etc. Há, no estilo, como em todos os fatos discursivos, um aspecto ligado à produção do texto e um relacionado a sua interpretação. Isto significa que o estilo toma forma na interação entre produção e interpretação, ou seja, numa práxis enunciativa, o que quer dizer que é um fato da ordem do acontecimento e não da estrutura.

[...]

O que determina um estilo é o conjunto de traços reiterados e não uma característica isolada. (FIORIN, 2000, p. 40).

De uma forma mais ampla, Martins (2012) se refere a estilo como tudo o que apresenta características particulares, observadas principalmente nas criações artísticas, não importando se são produções banais ou consideradas como mais requintadas. Martins também apresenta várias definições feitas por teóricos da área da Estilística sobre estilo, transcrevemos quatro delas que se completam.

Estilo é a qualidade do enunciado, resultante de uma escolha que faz, entre os elementos constitutivos de uma dada língua, aquele que a emprega em uma circunstância determinada. (MAROUZEAU *apud* MARTINS, 2012, p. 19)

Estilo é aspecto do enunciado que resulta de uma escolha dos meios de expressão, determinada pela natureza e pelas intenções do indivíduo que fala ou escreve. (GUIRAUD *apud* MARTINS, 2012, p. 19)

Estilo é o conjunto objetivo de características formais oferecidas por um texto como resultado da adaptação do instrumento linguístico às

finalidades do ato específico em que foi produzido. (HERCULANO DE CARVALHO *apud* MARTINS, 2012, p. 19)

Estilo é a linguagem que transcende do plano intelectual para carrear a emoção e a vontade. (MATTOSO CÂMARA *apud* MARTINS, 2012, p. 19)

Conforme Uchôa (2013), embora sejam muitas as definições sobre estilo, é interessante observar que elas se completam, abrangendo aspectos individuais e psíquicos, como escolha, elaboração e desvio da norma.

Citamos também Cardoso e Ignez (2013, p. 36) que afirmam que até mesmo em um conjunto de textos é possível perceber semelhanças ou diferenças de estilo. “Nesse caso, pode-se pensar em estilo como um uso determinado e marcado da língua, que se define por meio de escolhas de usos linguísticos operadas para a construção de discursos”.

Finalmente, corroboramos Morejón e Martins (2001) quanto ao estilo ser obtido graças às características da escrita de cada autor, refletindo sua história de vida e suas escolhas nos momentos de criação. Defendemos que não é possível produzir uma obra de cunho artístico sem que nela existam traços da personalidade de seu criador, assim como de sua história de vida. Estas marcas pessoais são motivadas por gostos, experiências vividas, emoções durante as criações e objetivos a serem obtidos pelas produções, entre outras causas. Por sua vez, entendemos que o autor perde o domínio de uma obra após a sua conclusão e divulgação, permitindo que seu estilo cause diversas reações e sentimentos junto ao público, podendo despertar desde o total desprezo até as mudanças mais radicais quanto às formas de viver, pensar e agir de cada pessoa.

Após esta introdução nos estudos sobre estilística, nos aprofundaremos em alguns itens importantes para nossa pesquisa, uma vez que a mesma aborda estilo lexical.

2.2.2 As teorias estilísticas de Charles Bally, Leo Spitzer e Amado Alonso

Diversos estudos consideram a Estilística proveniente de duas grandes correntes, sendo elas, a Estilística da Língua, de Charles Bally; e a Estilística Literária, de Leo Spitzer.

A Estilística da Língua, proposta por Charles Bally, volta-se para “os aspectos afetivos da língua falada, da língua a serviço da vida, língua viva, espontânea, mas gramaticalizada, lexicalizada, e possuidora de um sistema expressivo cuja descrição deve ser a tarefa da Estilística.” (MARTINS, 2012, p. 20).

Contrapondo as abordagens feitas por Martins (2012) e Viegas (1982), deduzimos que tais autores mencionam Bally como um defensor de que a língua deve ser um reflexo das atividades rotineiras e dos pensamentos individuais. Ela se constitui primeiramente pela racionalidade e impessoalidade, porém é modificada pela afetividade. Isto resulta na possibilidade de um mesmo tema ser expresso de diferentes formas, dependendo de quem o profere, das circunstâncias e do momento que ocorre.

Quando um indivíduo se encontra nas mesmas condições que os outros membros de seu grupo, podemos admitir a existência de uma norma que oriente e também determina seu comportamento na fala. No caso do escritor, já temos um emprego consciente da língua, além disso com deliberada intenção estética. Essa intenção é o que vai criar a distância imensa entre o falante do cotidiano e o falante artista. (VIEGAS, 1982, p. 168)

Viegas (1982) também atribui a Bally ter estabelecido à Estilística a “pesada tarefa do estudo dos fatos de linguagem, mas linguagem oral e organizada, organizada, ainda, do ponto de vista de seu conteúdo afetivo.” (VIEGAS, 1982, p. 168). A autora esclarece que, para Bally, o estilo está vinculado às escolhas e desvios provocados pelos sentimentos individuais.

Bally classifica como naturais os efeitos expressivos utilizados pelos indivíduos quando expõem seus sentimentos. De acordo com Martins (2012), neste grupo encontram-se manifestações como as de prazer, desprazer, admiração e desaprovação. Em outra esfera estão os efeitos evocativos que abarcam os meios sociais de quem os utiliza e da época e situações da produção das obras. Bally distancia-se do discurso individual, aproximando-se da língua coletiva, e defende que a Estilística deve estudar a linguagem em conjunto com a sensibilidade.

Bally inicia, assim, a Estilística da Língua ou da Expressão Linguística, que se ocupa da descrição do equipamento expressivo da língua como um todo, opondo a sua Estilística ao estudo dos estilos individuais e afastando-se, portanto, da literatura. (MARTINS, 2012, p. 21)

Cardoso e Ignez (2013) seguem o mesmo raciocínio de Martins (2012) e mencionam a Estilística de Bally vinculada à espontaneidade e à coletividade, sendo comum um falante fazer uso de palavras utilizadas pelas pessoas de sua comunidade linguística. Sendo assim, o estudo do estilo é possível graças a esta espontaneidade. Se por um lado os sujeitos falam de

forma comum, respeitando normas; por outro, cada pessoa usa as normas adequando-as ao seu modo peculiar de falar e às suas intenções.

Não se pode ignorar, como diz Bally, que existe um estilo na língua, mas também não se pode dizer que não existe estilo individual, na medida em que é o falante que escolhe como irá formular o seu enunciado. Isso promove uma dupla face ao estilo que permite sua manifestação na língua em sua natureza sistêmica e sua eclosão no discurso. (CARDOSO; IGNEZ, 2013, p. 41)

Citamos também Fiorin (2000) que explica que para Bally a linguagem conta com duas faces, uma afetiva e outra intelectual ou lógica. Isto é comprovado quando um mesmo conteúdo é expresso de maneiras diferentes, permitindo percebermos a lógica e a afetividade presentes no enunciado.

Outra corrente da Estilística foi proposta por Leo Spitzer. Motivado pela sua insatisfação diante da separação entre estudos linguísticos e literários, Spitzer concebe a Estilística Literária estruturada em três divisões, sendo elas, a idealista, de vertente filosófica; a psicológica, interessada na psicologia do autor; e a genética, que tenta chegar à origem da obra literária.

A Estilística de Spitzer parte da reflexão de cunho psicologista, sobre os desvios da linguagem em relação ao uso comum; uma emoção, uma alteração do estado psíquico normal, provoca um afastamento do uso linguístico normal; um desvio da linguagem usual é, pois, indício de um estado de espírito não habitual. O estilo do escritor – a sua maneira individual de expressar-se – reflete o seu mundo interior, a sua vivência. (MARTINS, 2012, p. 24)

De acordo com Morejón e Martins (2001), para Spitzer os fatores emotivos e psicológicos se sobrepõem, interferindo diretamente no modo como cada autor produz sua obra. Diante disso, Spitzer propôs o círculo filológico, método pelo qual o leitor, ao ler cuidadosamente uma obra e de forma intuitiva, percebe características estilísticas do autor. Tais características são uma espécie de pontapé inicial para estudos detalhados e se concluem em amplos entendimentos sobre o autor e sua obra. Spitzer considera o autor como o principal fator de toda obra artística e qualquer detalhe é útil quando percebido de uma forma intuitiva. “O salto inicial tem de ser dado mediante o uso da intuição e o ponto de partida pode ser dado perfeitamente por um traço linguístico individual.” (MOREJÓN; MARTINS, 2001, P. 161). Desta forma, de acordo com a Estilística Literária, de Spitzer, a intuição do leitor torna-se

indispensável e cada obra literária carrega o seu próprio universo imaginativo, afetivo e conceitual.

Percebe-se uma distinção entre as correntes de Charles Bally e Leo Spitzer, cuja composição foi analisada por Amado Alonso.

Segundo Amado Alonso (*apud* MARTINS, 2012), cabe à Estilística da Língua cuidar dos recursos expressivos linguísticos e valorativos e percebidos nas línguas falada ou literária, sendo ela a base da Estilística Literária da obra, com suas características particulares. Amado Alonso, defende que a tarefa da Estilística Literária é examinar como é constituída a obra literária, considerando o prazer estético provocado no leitor. Para ele, ao se fazer uma análise amparada na Estilística Literária, devem ser observados itens como dados históricos, sociológicos, ideológicos e psicológicos, além dos traços linguísticos. Tais características são capazes de revelar a visão de mundo do autor e seu prazer ao compor a obra, além daquele causado no leitor.

Ao abordar sobre Amado Alonso e sua proposta, Morejón e Martins (2001) pontuam que para muitos autores, ao se promover estudos estilísticos, deve-se desprezar valores que influenciam na produção da obra, como o ambiente de origem do autor, os fatos marcantes de sua vida, suas culturas e inspirações. Os autores mencionam que isto foi considerado um erro por Amado Alonso, que defendeu alguns fatores como essenciais para o bom entendimento da obra, sendo eles, as sensações, emoções, culturas, intuições e possibilidades expressivas.

Amado Alonso afirma que vistas assim as coisas, os conceitos complementares de fundo e forma reduzem-se a um conceito superior de forma, o que quer dizer que o estudioso opera apenas com os fatores a descoberto no texto, tal como ali eles estão dispostos por obra e graça da intuição e do pensamento do autor. Será, então, através da investigação do estilo, desta forma superior, que é bastante complexa que chegaremos a decifrar os múltiplos enigmas que a obra encerra e, finalmente, a ‘unicidade’ da mesma.” (MOREJÓN; MARTINS, 2001, p. 158).

Em nossa pesquisa, nos valem da proposta de Amado Alonso que, ao abordar o estilo, defende uma duplicidade que une Bally e Spitzer. Corroboramos Bally e sua Estilística da Língua, que promove uma união entre aspectos afetivos e normas gramaticais. Entretanto, também concordamos com Spitzer, pois o estilo é amplo e revela o artista e o sujeito em sua completude psicológica e ideológica por trás da obra, que também revela aspectos como a história de vida do autor e outros fatores pessoais que possam interferir na produção da obra.

Em seguida, abordaremos a teoria sobre a Linguística de Corpus, metodologia utilizada nesta pesquisa e que tornou possível desenvolvermos esta proposta de análises linguísticas de obras literárias.

2.3 LINGUÍSTICA DE *CORPUS*

A Linguística de *Corpus* (doravante LC) é uma metodologia que coleta e explora *corpora* com a finalidade de estudar a língua com o uso do computador. Segundo Berber Sardinha (2000), estas coletas são criteriosas e permitem extrema confiança nos resultados apresentados.

A Linguística de *Corpus* ocupa-se da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador. (BERBER SARDINHA, 2000, p. 325)

Já Beilke (2016) descreve a LC como uma abordagem metodológica de princípios descritivos, fundamentada em dados autênticos e que possibilita produzir saberes variados, “além de nos guiar a investigar hipóteses não pré-meditadas e permitir a descoberta e a comprovação de fatos linguísticos. (BEILKE, 2016, p. 72)

Com a LC é possível manipular um grande número de textos, não se atendo apenas a elaborar listas de palavras. Ela também facilita as pesquisas e as críticas literárias, “fornecendo entre outras possibilidades, palavras-chave dos textos estudados e linhas de concordâncias”. (GONÇALVES, 2008, p. 387)

Sobre a sua aplicação, Fromm (2003) apresenta uma lista de análises linguísticas possíveis com o uso da LC, entre elas a frequência de palavras e classes gramaticais mais comuns da língua, a regência dos verbos preposicionados, a seleção de uma nomenclatura para uma obra terminológica, a criação de dicionários gerais multilíngues, a base de dados para tradutores e o ensino de língua estrangeira.

Outros usos da LC são citados por Berber Sardinha (2000). O autor menciona que, além dos estudos da linguagem com a LC, destacam-se o seu uso em pesquisas linguísticas com fins comerciais e o desenvolvimento de técnicas com Processamento de Linguagem Natural (PLN), em conjunto com a Ciência da Computação. As possibilidades oferecidas pela LC também despertaram os interesses de universidades, editoras e empresas, como as de

comunicação, usuárias de *corpus* em atividades como processamentos automáticos de textos, informatizações de grandes bases de dados e montagens de sistemas de reconhecimento de voz. Com o uso da LC, pode-se concluir a verdade sobre hipóteses previamente elaboradas e ainda surgir outras deduções nunca pensadas.

Também ao se falar em LC, é comum surgirem questionamentos sobre o que exatamente é um *corpus*. Fromm (2003, p. 1) explica que “*corpus*, na área da Linguística, é uma coleção de textos reunidos, de áreas variadas ou não, com um propósito específico de análise”.

Sanchez (*apud* Berber Sardinha, 2000) define *corpus* como um conjunto de dados linguísticos orais ou escritos, produzidos segundo determinados critérios e de tal amplitude que permitam ser “representativos da totalidade do uso lingüístico ou de algum de seus âmbitos dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.” (SANCHEZ *apud* BERBER SARDINHA, 2000, p. 338).

A constituição de um *corpus* dever ser feita com textos naturais, configurando-se como “um artefato produzido para a pesquisa. Assim, se por um lado os textos devem ser naturais (autênticos e independentes do *corpus*), o *corpus* em si é artificial, um objeto criado com fins específicos de pesquisa”. (BERBER SARDINHA, 2000).

Ao se elaborar um *corpus*, é preciso ter clareza sobre como será a sua composição, seleção e função. Antes de iniciar é necessário que o seu idealizador já saiba o motivo da sua criação, pois uma produção aleatória corre o risco de se tonar sem utilidade.

Vale lembrar também que, embora o pesquisador não tenha calculado hipóteses teóricas sobre seu material de estudo na pesquisa dirigida/guiada pelo *corpus*, não podemos excluir totalmente a possibilidade de ele criar, ao menos, expectativas antes de dar início à sua investigação. Ademais, uma mesma pesquisa pode ser ao mesmo tempo dirigida/guiada pelo *corpus* e baseada em *corpus*, ou seja, é possível que o pesquisador trabalhe das duas formas ou que, de maneira imprevista, em algum momento sua pesquisa baseada em *corpus* seja guiada pelo *corpus*. (GRAMA, 2016, p. 107)

Conforme Shepherd (2012), quanto à tipologia e extensão do *corpus*, elas ficam a critério do pesquisador e de acordo com suas necessidades. Para algumas pesquisas é importante a utilização de um grande *corpus*, mas isso não é uma obrigatoriedade, pois *corpora* pequenos também podem revelar importantes hipóteses.

2.3.1 Um histórico da Linguística de *Corpus*

A LC surgiu diante da necessidade de realizar estudos com linguagem baseados em grandes *corpora*. Segundo Berber Sardinha (2000), durante a Antiguidade e a Idade Média já se produziam *corpora* com trechos da *Bíblia*. A prática se manteve e no início do século XX foi destinada principalmente ao ensino de línguas, ao contrário dos tempos atuais quando se prioriza a descrição da linguagem.

Antes do advento do computador, um histórico *corpus* foi produzido manualmente por Thorndike, em 1921, e contou com 4,5 milhões de palavras. Outro *corpus* de destaque foi o *Survey of English Usage (SEU)*, de 1953, também não computadorizado e planejado para ter um tamanho inicial de um milhão de palavras.

O Survey foi organizado em fichas de papel, cada uma contendo uma palavra do *corpus* inserida em 17 linhas de texto. As palavras foram analisadas gramaticalmente, com cada ficha recebendo uma categoria gramatical. O conjunto de categorias resultante serviu de base para o desenvolvimento dos etiquetadores computadorizados contemporâneos, que fazem a identificação de traços gramaticais automaticamente. (BERBER SARDINHA, 2000, p. 326)

Inicialmente os trabalhos desenvolvidos com base em *corpus* foram desmerecidos. Berber Sardinha (2000) relata que, ao final dos anos 50, Chomsky lançou *Syntactic Structures*, obra na qual criticou negativamente os trabalhos baseados em *corpus*. Na mesma época surgiram críticas e incertezas quanto à confiabilidade dos processamentos manuais de *corpora*. Dizia-se que o ser humano não era capaz de realizar este tipo de trabalho, mesmo quando era feito por grandes equipes.

A situação mudou com a criação do computador, principalmente no ano de 1964 com o lançamento do *Brown University Standard Corpus of Present-Day American English*, popularmente conhecido como *Corpus Brown*, composto por um milhão de palavras. Berber Sardinha (2000) destaca que foram utilizados cartões perfurados individualmente para transferir os textos para o computador. Este feito, por si só, já traria respeito e admiração à empreitada.” (BERBER SARDINHA, 2000, p. 324).

Tagnin (2018) destaca que o *corpus Brown* foi o responsável pela introdução dos estudos estatísticos da linguagem, precedendo a LC, e de modelo para outros *corpus*, como o *London-Oslo-Bergen Corpus (LOB Corpus)*, de 1978.

Quanto a referências de *corpora* eletrônicos, Berber Sardinha (2000) menciona o *Brown*, o *Birmingham University International Language Database (Birmingham Corpus)* e o *British National Corpus (BNC)*, todos em língua inglesa.

O *corpus Brown* é um marco por razões óbvias: é o pioneiro. O *corpus Birmingham* é importante porque foi o primeiro a ultrapassar a marca de 1 milhão de palavras iniciada pelo *Brown*. Vale lembrar que o *corpus Birmingham* se tornaria o *Bank of English*, sempre em crescimento, atingindo agora 320 milhões de palavras. Por fim, o *BNC* é um marco histórico porque foi o primeiro a conter 100 milhões de palavras e ainda é, dentre os *mega-corpora*, o único disponível para compra (dentro da Comunidade Européia apenas). (BERBER SARDINHA, 2000, p. 331)

Quanto a um paralelo da LC no mundo e no Brasil, segundo Biderman (1978), um dos estudos pioneiros utilizando *corpora* eletrônicos em língua portuguesa no Brasil foi o *Frequency Dictionary of Portuguese Words*. O material continha em torno de 500 mil palavras usadas no português europeu entre os anos 1920 e 1940.

Embora os primeiros *corpora* fossem compostos por poucos milhões de palavras, segundo Tagnin (2018), os atuais podem ultrapassar um bilhão de palavras, como o *News on te Web (NOW)* com mais de cinco bilhões de palavras, além do *Global Web-Based English (GloWbE)*, do *Wikipedia Corpus* e do *Corpus del Español*, sendo cada um composto por dois bilhões de palavras. Na língua portuguesa, a autora menciona o *Corpus do Português*, composto por 45 milhões de palavras do século XIII ao XX. Já Berber Sardinha (2000) destaca o *Corpus de Araraquara*; o *Corpus de Referência do Português Contemporâneo (CRPC)*, da Faculdade de Letras da Universidade de Lisboa; e o do Núcleo Interinstitucional de Linguística Computacional (doravante Nilc), sediado no Instituto de Ciências Matemáticas e de Computação (ICMC), da Universidade de São Paulo (doravante USP).

Em uma entrevista à Teixeira (2017), Tagnin destaca que, embora muitos pesquisadores ainda sejam relutantes em utilizar a LC em suas pesquisas, nos últimos anos ela foi difundida em vários segmentos no Brasil, como a Análise do Discurso. Tagnin também ressalta que é praticamente inimaginável fazer estudos em Terminologia sem usar a LC.

Com dados do ano de 2008, Berber Sardinha e Almeida (2008) apresentam um levantamento quanto a pesquisadores e grupos de pesquisa, entre outros itens, feito a partir de resultados disponibilizados por agências de fomento, como a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), do Ministério da Educação; e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (doravante CNPq), do

Ministério de Ciência, Tecnologia e Inovações. O levantamento revelou 132 nomes que em algum momento utilizaram a expressão Linguística de *Corpus* em seu currículo disponibilizado na Plataforma Lattes e 12 grupos de pesquisa em LC.

Ao realizarmos uma busca¹⁶ em 30 de novembro de 2021, pela Plataforma Lattes, do CNPq, na aba Currículo Lattes / Buscar Currículo, e inserir o assunto “Linguística de *Corpus*”, para doutores e demais pesquisadores, de nacionalidade brasileira, no modo de busca “Assunto (Título ou palavra-chave da produção)”, foram apresentados 2.579 resultados.

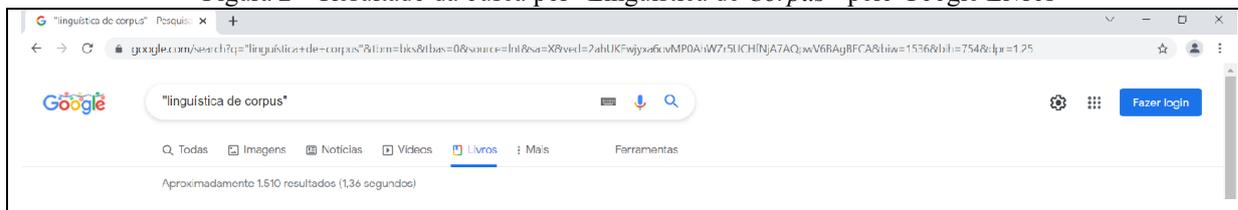
Figura 1: Resultado da busca por “Linguística de *Corpus*” pela Plataforma Lattes, do CNPq



Fonte: Plataforma Lattes, do CNPQ

Em outra pesquisa, também em 30 de novembro de 2021, utilizando o Google Livros¹⁷, ao buscar “Linguística de *Corpus*”, foram apresentados, aproximadamente, 1.510 resultados.

Figura 2 – Resultado da busca por “Linguística de *Corpus*” pelo Google Livros



Fonte: Google Livros

Novamente em 30 de novembro de 2021, o Google Acadêmico¹⁸ apresentou 4.670 resultados para “Linguística de *Corpus*” nas páginas em português.

Figura 3 – Resultado da busca por “Linguística de *Corpus*” pelo Google Acadêmico

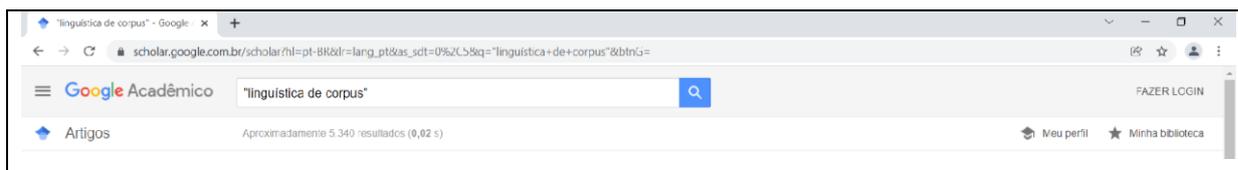
¹⁶ Disponível em: <<http://buscatextual.cnpq.br/buscatextual/busca.do>>. Acesso em 30 nov 2021.

¹⁷ Disponível em:

<<https://www.google.com/search?q=%22lingu%C3%ADstica+de+corpus%22&tbm=bks&tbas=0&source=Int&sa=X&ved=2ahUKEwjyxa6ovMP0AhWZr5UCHfNjA7AQpwV6BAgBECA&biw=1536&bih=754&dpr=1.25>>. Acesso em 30 nov 2021.

¹⁸ Disponível em: <[https://scholar.google.com.br/scholar?hl=pt-](https://scholar.google.com.br/scholar?hl=pt-BR&lr=lang_pt&as_sdt=0%2C5&q=%22lingu%C3%ADstica+de+corpus%22&btnG=>)

[BR&lr=lang_pt&as_sdt=0%2C5&q=%22lingu%C3%ADstica+de+corpus%22&btnG=>](https://scholar.google.com.br/scholar?hl=pt-BR&lr=lang_pt&as_sdt=0%2C5&q=%22lingu%C3%ADstica+de+corpus%22&btnG=>). Acesso em 30 nov 2021.



Fonte: Google Acadêmico

Comparando números apresentados em anos anteriores e em 2021, concordamos com Novodvorski e Finatto (2014) quando afirmam ter ficado claro que a LC vai muito além de “contar palavras” (NOVODVORSKI; FINATTO, 2014, p. 15), oferecendo importantes contribuições para comunidades de pesquisas nacionais e internacionais.

2.4 ESTILÍSTICA DE *CORPUS*

Outra teoria da qual nos apropriamos é a Estilística de *Corpus* (doravante EC). Devido a ser algo relativamente novo, o segmento conta com poucos materiais a seu respeito se comparado aos demais já citados neste capítulo.

Mahlberg (2020) define a EC como um cruzamento entre os estudos da linguagem e da literatura, útil em ambas as áreas e possibilitado pelo uso de ferramentas descritivas disponibilizadas pela LC. O uso em conjunto da Estilística Literária com a LC, resultando a EC, permite mensurar a criatividade presente em um texto. Ressaltamos que é indispensável conhecer a Estilística e a LC para utilizar a EC.

O que a Estilística de *Corpus* pode fazer além da óbvia provisão de dados quantitativos é auxiliar a análise de um texto individual, fornecendo várias opções para a comparação de um texto com grupos de outros textos para identificar tendências, relações intertextuais ou reflexões de contextos social e cultural. (MAHLBERG, 2020, p. 4432).

Assim como a LC, inicialmente a EC foi recebida com reservas pelos estudiosos. Segundo Shepherd e Berber Sardinha (2013), os dados gerados pela EC foram considerados descontextualizados e classificados como reproduções de trechos dos textos originais. Segundo os críticos, isto praticamente impedia análises estilísticas mais precisas. Eles defendiam que tal processo precisava de leituras atentas de todo o texto e não só de alguns trechos. Com o tempo, o cenário se alterou e a EC deixou de ser criticada, firmando-se como uma área promissora por possibilitar melhores análises dos traços estilísticos, identificar com mais precisão as caracterizações e distinções de cada texto e permitir percepções, como de um

palavreado específico e, até mesmo, de traços de criatividade ou de repetições. Podemos citar, como exemplo, que a EC revela características distintas em falas de personagens, escritas dos autores e obras ou conjunto delas, além do uso de adjetivos, substantivos e interjeições. Os estudos com a EC se mostram ainda de grande utilidade em análises como as de padrões de discurso, mudanças históricas e variações lexicais, gramaticais e de registros falados e escritos.

Conforme Shepherd (2004, p.21), ao fazer uso da EC é possível observar o raro, o que nem sempre é facilmente percebido apenas pela análise humana. “(...) o computador também nos ajuda a observar o raro, o marginal, o atípico e o caso único. Ao se compararem dois *corpora*, mostra-se o que é característico de cada um deles”. Em relação às principais respostas apresentadas pelo uso da EC, a autora menciona itens como estilos individuais dos autores e das obras, estabelecimentos de autorias e características de um gênero específico ou de um período literário.

Sobre a prática das análises, Mahlberg (2020) aponta que primeiramente deve-se analisar palavras-chave, principalmente os substantivos próprios e as palavras de conteúdo. Quanto a isto, Shepherd e Berber Sardinha (2013) destacam que as palavras-chave são indicadores úteis quanto ao estilo do *corpus* analisado, inclusive sobre seu enredo.

Entretanto, Burrows (*apud* Mahlberg, 2020) defende que não se pode fazer estudos baseados apenas em palavras-chave ou de conteúdo, pois excluir palavras funcionais, como artigos e conjunções, pode acarretar análises incompletas. Burrows aponta que as palavras ocorrem simultaneamente, em uma sequência lógica de enunciação, sendo assim, as frequências devem ser analisadas por completo.

Em relação a esta divergência quanto ao uso ou não apenas de palavras-chaves nos estudos envolvendo a EC, nos posicionamos a favor de ambas as teorias, pois é preciso analisar qual é o método mais útil para ser empregado em cada pesquisa.

Sobre agrupamentos lexicais, Mahlberg (2020) considera que os mesmos também devem ser considerados e utilizados para se entender mais detalhadamente os significados de um texto, identificar funções textuais e perceber traços de uma linguagem mais ou menos culta, palavras típicas de cada personagem, perfis sociais e culturais, enfim, uma infinidade de opções possíveis. Destacamos que, segundo a autora, os agrupamentos lexicais longos são mais adequados para se observar uma obra, enquanto que os menores, geralmente, evidenciam tendências gerais presentes em vários textos.

Ao se utilizar a EC deve-se observar o que se destaca e se desvia das normas linguísticas, causando efeitos estilísticos e psicológicos nos leitores, embora nem todo efeito especial em um texto possa ser considerado como característica de estilo. “A multifuncionalidade das palavras e a flexibilidade com a qual as palavras podem entrar em relações textuais com outras palavras possibilitam vários efeitos linguísticos.” (MAHLBERG, 2020, p. 4.434).

Mahlberg (2020) destaca que, diante das pesquisas com a EC, o computador se tornou equipamento indispensável para se obter levantamentos confiáveis. A sua utilização não acarreta desprezo com os estudos anteriores e métodos convencionais, como alguns estudiosos acreditam. Pelo contrário, a confiabilidade do equipamento é constatada ao se observar a semelhança entre os dados obtidos por métodos tradicionais e manuais.

O ponto importante é que observações da Linguística de *Corpus* não começam com tantas restrições determinadas por pressupostos anteriores de modo que a Estilística de *Corpus* pode contar com o potencial da Linguística de *Corpus* e da Estilística Literária de se complementarem mutuamente. (MAHLBERG, 2020, p. 4438).

O uso do computador e o aumento da quantidade de materiais e *corpora* eletrônicos disponíveis e utilizados em análises interpretativas favoreceram a EC, segundo Teixeira e Ottaiano (2017). Os autores reforçam que isto ocorre devido às escolhas dos pesquisadores por métodos capazes de realizar análises mais objetivas, tornando a EC um dos principais recursos disponíveis na atualidade para investigações linguístico-literárias.

2.4.1 Estilística de *Corpus* e complexidade textual

Defendemos que uma das análises que a EC permite diz respeito à complexidade textual (doravante CT) de um texto, como as obras literárias. Um dos pioneiros a estudar a CT e a desenvolver fórmulas para estimá-la foi Rudolph Flesch, criador do índice Flesch, Segundo Finatto (2011), o índice foi criado nos EUA de acordo com a realidade educacional do país, a partir de percepções de Rudolph Flesch diante das dificuldades dele e de seus companheiros de fábrica em ler manuais escritos em inglês.

Finatto (2011) explica que o índice Flesch agrupa métricas e calcula coesão, coerência e nível de CT, promovendo uma correlação entre tamanhos médios de palavras e sentenças.

De acordo com a formatação do índice Flesch¹⁹, quanto maior o resultado da métrica, menor a CT e mais fácil é a compreensão. Ao contrário, quanto menor o resultado da métrica, maior é a CT e mais difícil é a compreensão.

Os cálculos do índice Flesch podem ser feitos pela página do Núcleo Interinstitucional de Linguística Computacional Metrix (NILC-Metrix), elaborada pelo Departamento de Ciência da Computação e do Instituto de Física, da USP²⁰. Entretanto, conforme ressalta Paraguassu (2018), o índice Flesch não poder ser considerado determinante absoluto do grau de dificuldade de um texto, mas pode ser um indicativo nas análises que contemplam as possibilidades de compreensão de um leitor ou grupo de leitores. Destacando que para uma pessoa um texto pode ser de fácil compreensão, enquanto que para outra pode ser de difícil inteligibilidade.

¹⁹ Disponível em http://fw.nilc.icmc.usp.br:23380/metrixdoc#indices_leiturabilidade. Acesso em 4 jul. 2021.

²⁰ Disponível em <http://fw.nilc.icmc.usp.br:23380/metrixdoc#flesch>. Acesso em 4 jul. 2021.

3 METODOLOGIA DE PESQUISA

Neste capítulo apresentamos as etapas e os procedimentos metodológicos aplicados em nossa pesquisa e montagem dos *corpora*.

3.1 Preparação e montagem do *corpora*

A preparação dos *corpora* da pesquisa teve início no mês de agosto de 2019, sendo dividido entre romances literários de Paulo Coelho e obras famosas da literatura brasileira e de diversos autores.

3.1.1 *Corpus* das obras de Paulo Coelho

Inicialmente, durante o segundo semestre de 2019, adquirimos os 16 romances literários de Paulo Coelho. Analisamos somente os romances pois foram as obras fundamentais pela projeção do nome do autor. Destacamos que a escolha não diminui a importância de suas demais produções, como composições musicais e textos jornalísticos.

Montamos o *corpus* somente com o texto principal de cada romance e os prólogos, pois em algumas obras esta parte diz respeito ao enredo. Excluímos notas de rodapé e do autor, dedicatórias e demais citações, que não interferem no enredo dos romances e não se configuram como um processo espontâneo de produção da linguagem.

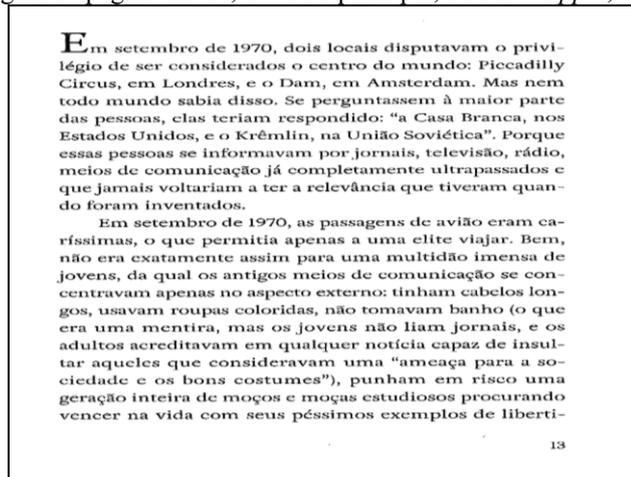
Os romances são, em ordem de lançamento e com as respectivas editoras, *O diário de um mago* (1987, Rocco), *O alquimista* (1988, Rocco), *Brida* (1990, Rocco), *As Valkírias* (1992, Rocco), *Na margem do rio Piedra eu sentei e chorei* (1994, Rocco), *O Monte Cinco* (1996, Objetiva), *Verônica decide morrer* (1998, Objetiva), *O demônio e a srta. Prym* (2000, Objetiva), *Onze minutos* (2003, Rocco), *O Zahir* (2005, Rocco), *A bruxa de Portobello* (2006, Planeta), *O vencedor está só* (2008, Agir), *O Aleph* (2010, Sextante), *Adultério* (2014, Sextante), *A espiã* (2016, Paralela) e *Hippie* (2018, Paralela).

Em 2019 tais obras não estavam disponíveis para domínio público e *downloads* pela internet. Optamos por utilizar as edições impressas e mencionar quais editoras as lançaram, dando as devidas atribuições quanto aos direitos autorais. Utilizamos edições dos anos de lançamento das obras por serem mais fidedignas ao estilo de escrita de Paulo Coelho e que

não tiveram alterações e correções, inclusive devido ao Acordo Ortográfico da Língua Portuguesa de 1990.

Ao iniciarmos as digitalizações dos livros impressos, encadernados com lombadas quadradas e com as páginas juntas, precisamos recortar as lombadas, deixando as folhas soltas, e digitalizamos cada romance, página por página, na extensão PDF.

Figura 4: Imagem da página inicial, do texto principal, da obra *Hippie*, de Paulo Coelho



Fonte: Elaboração própria

Em seguida, convertemos cada arquivo em texto do Word, formato doc, utilizando endereços eletrônicos (*sites*), como o ILovePDF²¹ e o Smallpdf²², que disponibilizavam a tecnologia *Optical Character Recognition* (doravante OCR), capaz de reconhecer e tornar editáveis caracteres de textos impressos e imagens digitalizadas, datilografadas ou escritas à mão. Para cada romance foi criado um arquivo em Word.

No formato Word, procedemos com as conferências manuais entre textos digitais e impressos e corrigimos diversas alterações em palavras feitas automaticamente pelo conversor. Em todas as 16 obras convertidas ocorreram também omissões de trechos e de parágrafos inteiros, que nos forçou a digitar o conteúdo.

A seguir citamos exemplos das alterações feitas pelo corretor durante as conversões e as correções manuais feitas por nós para ficarem iguais às versões impressas.

Quadro 1 – Obra *O Aleph* – Exclusões e correções das sílabas “fi”

Na obra <i>O Aleph</i> todas as sílabas “fi” foram apagadas	
Versão após conversão, na qual pode se observar as ausências da sílaba “fi”.	Versão após conferências e correções das palavras “justifica” e “ficam” em

²¹ Disponível em: https://www.ilovepdf.com/pt/pdf_para_word. Primeiro acesso em 20 set 2019.

²² Disponível em: <https://smallpdf.com/pt/pdf-para-word>. Primeiro acesso em 20 set 2019.

	conformidade com a versão impressa de 2010, p. 12, 1º §.
(...) a nota que justi ca uma sinfonia inteira, a letra que resume o livro. Passo por um período de euforia, que aos poucos vai desaparecendo. Algumas coisas cam para sempre (...)	(...), a nota que justifica uma sinfonia inteira, a letra que resume o livro. Passo por um período de euforia, que aos poucos vai desaparecendo. Algumas coisas ficam para sempre (...)

Fonte: Elaboração própria.

Quadro 2 – Obra *O Aleph* – Supressões e correções das junções das letras “fl”

Em <i>O Aleph</i> foram apagadas as junções das letras “fl”	
Versão após conversão, na qual pode se observar a ausência da junção das letras “fl”	Versão após conferência e correção da palavra “florista” em conformidade com a versão impressa de 2010, p. 252, 4º §.
<i>No caminho, passo pela orista e compro uma dúzia de rosas.</i>	<i>No caminho, passo pela florista e compro uma dúzia de rosas.</i>

Fonte: Elaboração própria.

Quadro 3 – Obra *Hippie* – Alterações e correções do nome “Karla”

Em <i>Hippie</i> houve alterações como o nome “Karla” que foi alterado para “<arla”	
Versão após conversão, na qual pode se observar “I<arla”	Versão após conferência e correção do nome “Karla” em conformidade com a versão impressa de 2018, p. 140, 1º §.
<i>Quanto mais Rayan se interessava por I<arla, mais Mirthe ficava insegura, furiosa (...)</i>	<i>Quanto mais Rayan se interessava por Karla, mais Mirthe ficava insegura, furiosa (...)</i>

Fonte: Elaboração própria.

Quadro 4 – Obra “Hippie” – Conversões de aspas na sequência “ii”

Em <i>Hippie</i> algumas aspas foram convertidas em “ii”,	
Versão após conversão, na qual pode se observar “cedo?ii”	Versão após conferência e correção de “cedo?” em conformidade com a versão impressa de 2018, p. 122, 9º §
<i>"Mas não é muito cedo?ii</i>	<i>"Mas não é muito cedo?"</i>

Fonte: Elaboração própria.

Quadro 5 – Obra *Hippie* – Alterações da letra “m” para a junção das letras “rn”

Em <i>Hippie</i> ocorreu a alteração da letra “m” para a junção das letras “rn”,	
Versão após conversão, na qual pode se observar “urna” e “corno”	Versão após conferências e correções de “uma” e “como” em conformidade com a versão impressa de 2018, p. 171, 4º §.
<i>Quando chegou à Cidade do Cabo, resolveu descansar antes de procurar urna ordem religiosa e oferecer-se corno probacionista.</i>	<i>Quando chegou à Cidade do Cabo, resolveu descansar antes de procurar uma ordem religiosa e oferecer-se como probacionista.</i>

Fonte: Elaboração própria.

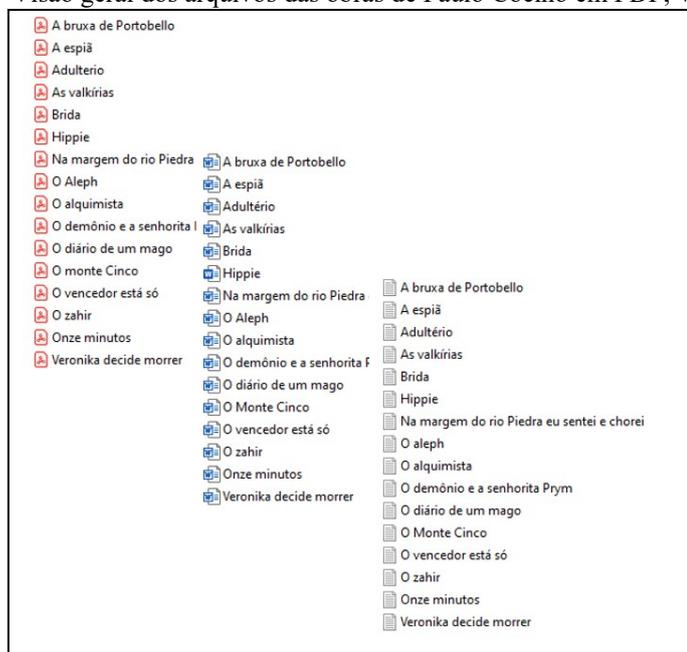
Quadro 6 – Obra *O vencedor está só* – Correções das hifenizações

As hifenizações das versões impressas foram mantidas após as digitalizações e conversões de PDF para Word, causando erros de acordo com as obras impressas e que foram corrigidos manualmente, como neste trecho de <i>O vencedor está só</i>.	
Versão após conversão, na qual pode se observar “mi- lionários”, “namo- rada” e “cor- re”.	Versão após conferências e correções “milionários”, “namorada” e “corre”, em conformidade com a versão impressa de 2008, p. 20, 6º §.
(...) deixam Igor em paz, pensando que deve ser um desses mi- lionários que vão a Cannes apenas para ver se encontram uma namo- rada . (...) o boato cor- re , todos ali já sabem que o homem solitário não pertence à indústria do cinema ou da moda (...)	(...) deixam Igor em paz, pensando que deve ser um desses milionários que vão a Cannes apenas para ver se encontram uma namorada . (...) o boato corre , todos ali já sabem que o homem solitário não pertence à indústria do cinema ou da moda (...)

Fonte: Elaboração própria.

Após corrigir manualmente as obras no formato Word, conferimos todas pelo recurso “Revisão de texto”, do programa Microsoft Word. A última etapa da preparação do *corpus* foram as conversões dos textos em Word para o formato TXT (Bloco de notas), na codificação ANSI, necessárias para o melhor processamento no programa *WordSmith Tools*. Criamos outra pasta no recurso “Explorador de arquivos” para os arquivos no formato TXT.

Figura 5 – Visão geral dos arquivos das obras de Paulo Coelho em PDF, Word e TXT



Fonte: Elaboração própria.

A seguir apresentamos um quadro com as atividades feitas durante a montagem do *corpus* com as obras de Paulo Coelho e do tempo empregado em cada tarefa.

Quadro 7 – Montagem do *corpus* do Paulo Coelho – Atividades realizadas e tempo empregado

ATIVIDADES REALIZADAS	TEMPO EMPREGADO
Aquisição das obras literárias	Seis meses de buscas em sebos literários e internet
Total de obras no <i>corpus</i>	16 obras
Digitalização das obras	10 dias com uma média de duas horas diárias de atividades
Conversões de PDF para Word	8 dias com uma média de uma hora diária de atividades
Conferências e correções em cada obra	10 dias em cada obra com uma média de quatro horas diárias de atividades
Conferências e correções em todas as obras	160 dias, 22 semanas, 640 horas
Páginas impressas conferidas	3.980 páginas
Conferências de cada obra no recurso “Revisão de texto”	3 em cada obra com uma média de uma hora para cada obra
Conferências totais nas 16 obras	48 conferências
Conversões de Word para TXT	1 dia com uma média de duas horas de atividades
Total de horas trabalhadas	686 horas

Fonte: Elaboração própria

3.1.2 *Corpus* das obras dos demais autores

Montamos um segundo *corpus* com 17 obras famosas de autores diversos e pertencentes a diferentes escolas literárias. Assim como no *corpus* de Paulo Coelho, a metodologia utilizada foi a da LC e o *corpus* é composto pelos textos principais de prosas, poemas, cartas e peças teatrais, tendo sido excluídas as notas de rodapé, notas do autor, dedicatórias e demais citações. Também retiramos textos inteiramente escritos em outros idiomas para manter o mesmo padrão das obras de Paulo Coelho utilizadas na pesquisa e produzidas em português.

Fizemos uso de 13 obras já disponíveis em domínio público, sendo elas em ordem de escolas literárias e com os respectivos anos de lançamento: *A carta*, de Pero Vaz de Caminha (Literatura de Informação, 1500); *Auto representado na Festa de São Lourenço*, de Padre José de Anchieta (Literatura de Catequese, 1583); *Sermão de Santo Antônio aos peixes*, de Padre Antônio Vieira (Barroco, 1682); *Seleção de obras poéticas*, de Gregório de Matos (Barroco, antologia, diversos anos); *Marília de Dirceu*, de Tomás Antônio Gonzaga (Arcadismo, 1792); *Iracema*, de José de Alencar (Romantismo, 1ª fase, 1865); *Lira dos vinte anos*, de Álvares de Azevedo (Romantismo, 2ª fase, 1853); *Espumas flutuantes*, de Castro Alves (Romantismo, 3ª fase, 1870); *Dom Casmurro*, de Machado de Assis (Realismo, 1899); *O cortiço*, de Aluísio Azevedo (Naturalismo, 1890); *Alma inquieta*, de Olavo Bilac (Parnasianismo, antologia,

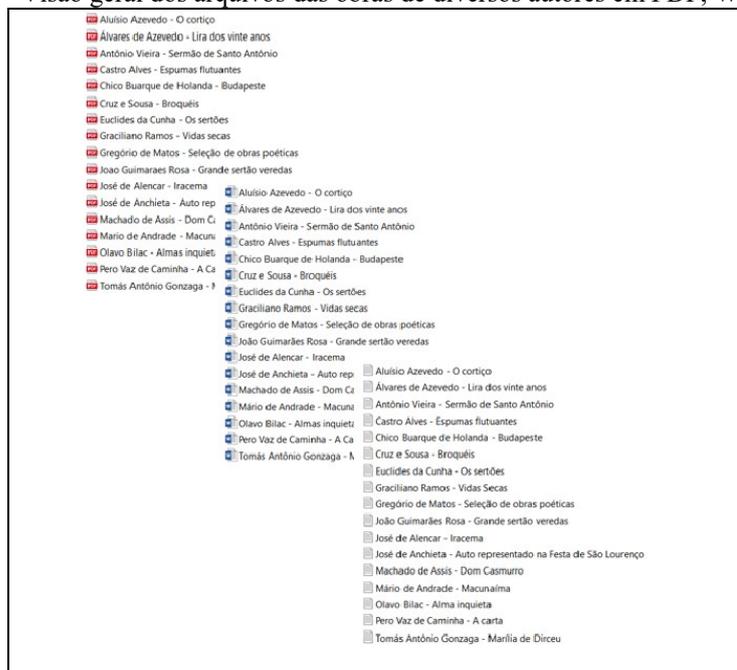
diversos anos); *Broquéis*, de Cruz e Sousa (Simbolismo, 1893) e *Os sertões*, de Euclides da Cunha (Pré-Modernismo, 1902).

Fizemos os *downloads* das obras no formato PDF e criamos arquivos para cada uma em pasta específica. Posteriormente convertemos cada obra em texto do Word, extensão doc, pelos mesmos endereços eletrônicos (*sites*) utilizados com as obras de Paulo Coelho. Nesta conversão para o Word, cada linha do texto em PDF, foi transformada em um parágrafo. Tivemos que realizar conferências entres Word e PDF e as correções no Word.

Quanto às demais quatro obras utilizadas, são elas com respectivos títulos, autores, escolas literárias, anos de lançamentos e editoras: *Macunaíma: o herói sem nenhum caráter*, de Mário de Andrade (Modernismo, 1ª fase, 1928, Villa Rica), *Vidas secas*, de Graciliano Ramos (Modernismo, 2ª fase, 1938, Record), *Grande sertão: veredas*, de Guimarães Rosa (Modernismo, 3ª fase, 1956, Companhia de Bolso) e *Budapeste*, de Chico Buarque (Contemporânea, 2003, Companhia das Letras).

Como tais obras não se encontravam em domínio público em julho de 2021, procedemos como nas obras de Paulo Coelho quanto às aquisições, digitalizações, conversões, conferências e correções. Outra pasta foi criada para abrigar os arquivos no formato TXT.

Figura 6 – Visão geral dos arquivos das obras de diversos autores em PDF, Word e TXT



Fonte: Elaboração própria.

A seguir apresentamos um quadro com as atividades feitas durante a montagem do *corpus* com as obras dos demais autores e do tempo empregado em cada tarefa.

Quadro 8 – Montagem do *corpus* dos demais autores – Atividades realizadas e tempo empregado

ATIVIDADES REALIZADAS	TEMPO EMPREGADO
Total de obras de diversos autores	17 obras
Total de obras por <i>downloads</i>	13 obras
<i>Downloads</i> das obras	2 dias com média diária de duas horas de atividades
Total de obras impressas	4 obras
Aquisição das obras literárias impressas	1 dia
Digitalização das obras impressas	4 dias com média diária de uma hora de atividades
Conversões de PDF para Word das obras digitais ou digitalizadas	1 dia com média diária de uma hora de atividades
Conferências e correções em cada obra (digitais e impressas)	4 dias em cada uma das 13 obras obtidas por <i>downloads</i> e em cada uma das quatro obras adquiridas no formato impresso com média de cinco horas diárias de atividades
Páginas impressas conferidas	911 páginas
Conferências e correções em todas as obras	68 dias, 10 semanas, 340 horas
Conferências de cada obra no recurso “Revisão de texto”	3 em cada obra com média de uma hora para cada obra
Conferências totais nas 17 obras	39 conferências
Total de horas trabalhadas	366 horas

Fonte: Elaboração própria

3.1.3 Usando o *WordSmith Tools*

Em nossas análises exploramos o *corpora* das obras de Paulo Coelho e dos demais autores tendo como ferramentas de pesquisa no formato digital o programa *WordSmith Tools* (doravante WST), versão 6.0, de Scott (2012), e o índice Flesch.

Berber Sardinha (1999, 2009) descreve o WST como um conjunto de programas integrados, útil em pesquisas de LC e com recursos capazes de analisar frequências e coocorrências das palavras constantes nos *corpora*. O WST conta com as ferramentas *WordList*, *Keywords* e *Concord*, que possibilitam observar aspectos da linguagem através de análises probabilísticas lexicais.

Inicialmente utilizamos a ferramenta *WordList* que, segundo Berber Sardinha (1999), oferece listas de palavras, sendo capaz de produzir simultaneamente duas listas, uma em ordem alfabética e outra com as palavras em ordem de frequência. Utilizamos as *wordlists* para obtermos listas de palavras das obras analisadas com total de *tokens*, *types* e *standardised*

type/token ratio (doravante STTR). Pelo item *tokens* temos o total geral de ocorrências de palavras em cada obra, considerando as repetições da primeira à última. Em *types* percebemos o total de palavras em cada obra sem considerar as repetições. Em STTR consta uma relação padronizada obtida pela divisão do total de *types* pelo total de *tokens*, multiplicado por 100, para se obter os valores em porcentagem e termos conhecimento da densidade lexical de cada obra. Segundo Silva (2009), a STTR determina a densidade lexical de uma obra ao calcular a porcentagem de tipos de palavras diante do total de palavras do texto e quanto mais elevada a STTR, maior é a densidade lexical utilizada.

Nesta fase, produzimos duas *wordlists* para cada obra, sendo uma sem *stoplist* e outra com uma *stoplist* fornecida pelo orientador e de autoria da doutora em Estudos Linguísticos, pela UFU, Flávia Santos da Silva. Esta última conta com 995 palavras gramaticais do português que não tínhamos interesse de analisar, sendo elas pronomes, artigos, numerais, preposições, conjunções e interjeições. Não constam na *stoplist* as palavras lexicais que temos interesse de analisar, sendo elas substantivos e adjetivos. Para cada romance literário de Paulo Coelho e para o conjunto de todos eles criamos uma *wordlist*. Utilizamos o mesmo procedimento para as obras dos demais autores incluídos na pesquisa.

Figura 7: *Wordlist*, com *stoplist*, com as 20 primeiras palavras de todas as obras de Paulo Coelho



N	Word	Freq.	%	Texts	%	Lemmas	Set
1	NÃO	13.866	1,60	16	100,00		
2	É	5.431	0,63	16	100,00		
3	ERA	4.229	0,49	16	100,00		
4	ESTAVA	3.745	0,43	16	100,00		
5	VOCÊ	3.382	0,39	16	100,00		
6	ESTÁ	2.873	0,33	16	100,00		
7	QUANDO	2.743	0,32	16	100,00		
8	JÁ	2.631	0,30	16	100,00		
9	TINHA	2.397	0,28	16	100,00		
10	DISSE	2.296	0,26	16	100,00		
11	SER	2.179	0,25	16	100,00		
12	VIDA	2.104	0,24	16	100,00		
13	HAVIA	2.087	0,24	16	100,00		
14	ATÉ	2.015	0,23	16	100,00		
15	MUNDO	1.902	0,22	16	100,00		
16	TEMPO	1.796	0,21	16	100,00		
17	FOI	1.773	0,20	16	100,00		
18	PESSOAS	1.649	0,19	16	100,00		
19	VEZ	1.597	0,18	16	100,00		
20	DIA	1.561	0,18	16	100,00		

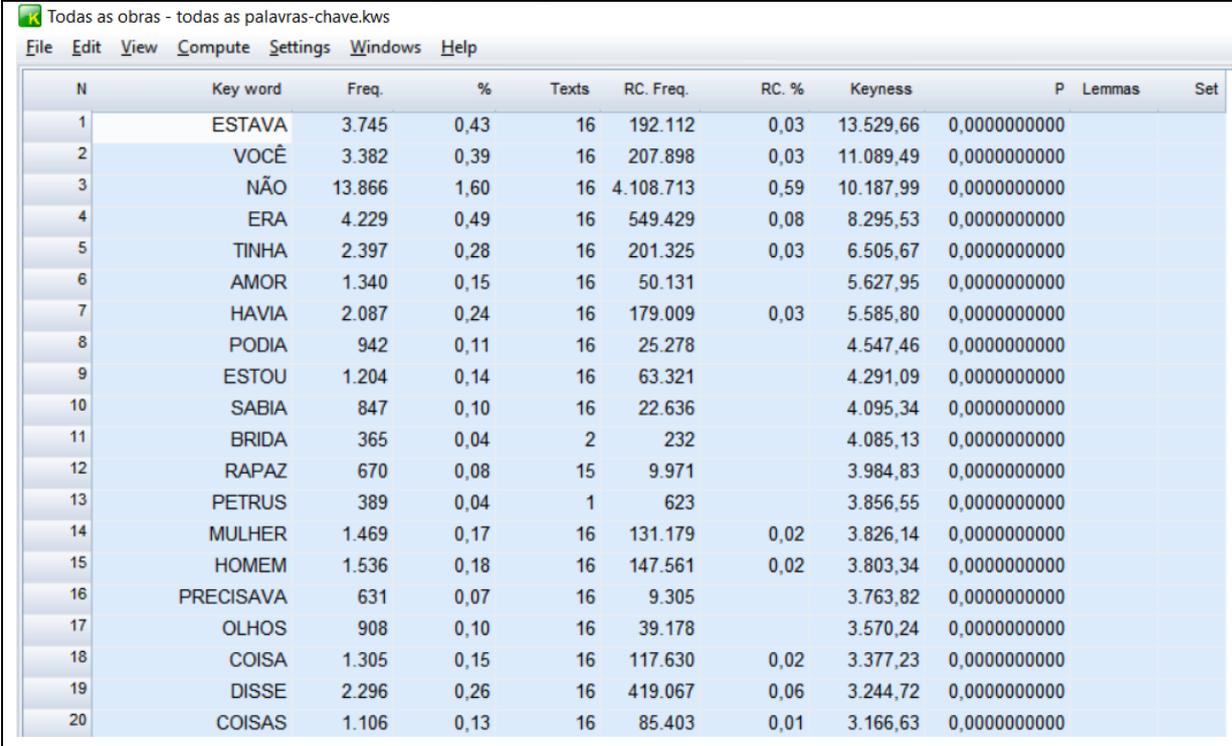
Fonte: Elaboração própria.

Com a *wordlist* sem *stoplist* tivemos acesso não somente às palavras mais frequentes, mas também às *hapax legomena*. Conforme descrevem Fromm, Santos, Grama e Beilke (2020), o WST oferece tal lista por meio de um procedimento simples com a utilização apenas da barra de rolagem. Com o resultado podemos analisar a densidade lexical de um texto, *corpus* ou *corpora*.

Em seguida utilizamos o programa *KeyWords*, do WST, definido por Berber Sardinha (2009) como capaz de produzir listas de palavras-chave. O *KeyWords* permitiu selecionarmos itens de uma lista de palavras e compararmos suas frequências com a lista de referência BP 100K. A lista BP 100K foi cedida pelo orientador e escolhida devido ao seu conteúdo com 100 mil primeiras palavras oriundas do *Corpus Brasileiro (Brazilian Corpus)*, um dos maiores acervos da língua portuguesa brasileira e disponibilizado *on-line* pelo Centro de Pesquisas, Recursos e Informação de Linguagem (Cepril), do Programa de Pós-Graduação em Linguística Aplicada (Lael), da Pontifícia Universidade Católica de São Paulo (PUCSP)²³.

Esta etapa foi realizada para analisarmos as palavras-chave e os campos lexicais do conjunto das obras de Paulo Coelho. No WST, a chavicidade das palavras no texto ou conjunto de textos é apresentada pela coluna *Keyness*.

Figura 8: *KeyWord* com as 20 primeiras palavras-chave dos romances literários de Paulo Coelho



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P Lemmas	Set
1	ESTAVA	3.745	0,43	16	192.112	0,03	13.529,66	0,0000000000	
2	VOCÊ	3.382	0,39	16	207.898	0,03	11.089,49	0,0000000000	
3	NÃO	13.866	1,60	16	4.108.713	0,59	10.187,99	0,0000000000	
4	ERA	4.229	0,49	16	549.429	0,08	8.295,53	0,0000000000	
5	TINHA	2.397	0,28	16	201.325	0,03	6.505,67	0,0000000000	
6	AMOR	1.340	0,15	16	50.131		5.627,95	0,0000000000	
7	HAVIA	2.087	0,24	16	179.009	0,03	5.585,80	0,0000000000	
8	PODIA	942	0,11	16	25.278		4.547,46	0,0000000000	
9	ESTOU	1.204	0,14	16	63.321		4.291,09	0,0000000000	
10	SABIA	847	0,10	16	22.636		4.095,34	0,0000000000	
11	BRIDA	365	0,04	2	232		4.085,13	0,0000000000	
12	RAPAZ	670	0,08	15	9.971		3.984,83	0,0000000000	
13	PETRUS	389	0,04	1	623		3.856,55	0,0000000000	
14	MULHER	1.469	0,17	16	131.179	0,02	3.826,14	0,0000000000	
15	HOMEM	1.536	0,18	16	147.561	0,02	3.803,34	0,0000000000	
16	PRECISAVA	631	0,07	16	9.305		3.763,82	0,0000000000	
17	OLHOS	908	0,10	16	39.178		3.570,24	0,0000000000	
18	COISA	1.305	0,15	16	117.630	0,02	3.377,23	0,0000000000	
19	DISSE	2.296	0,26	16	419.067	0,06	3.244,72	0,0000000000	
20	COISAS	1.106	0,13	16	85.403	0,01	3.166,63	0,0000000000	

Fonte: Elaboração própria

²³ Disponível em <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>. Acesso em 19 nov 2021.

A terceira etapa de uso do WST consistiu em utilizar o programa *Concord* que nos ofereceu as linhas de concordâncias das palavras-chave de Paulo Coelho analisadas na pesquisa. Tais concordâncias são listas de ocorrências de palavras de busca vinculadas ao texto que os acompanham, também chamados cotexto, conforme Berber Sardinha (1999).

Criamos duas listas com as palavras *Deus* e *demônio*, utilizadas no conjunto dos 16 romances de Paulo Coelho. Segundo o *Michaelis Dicionário Brasileiro de Língua Portuguesa*²⁴, *Deus*, com inicial maiúscula, é “o ser supremo; o espírito infinito e eterno, criador e preservador do Universo; incriado. Ente tríplice e uno, infinitamente perfeito, livre e inteligente, criador e regulador do Universo”. Já a definição de *demônio* apresentada pelo mesmo dicionário, segundo as religiões cristãs, é “o anjo que se rebelou contra Deus e que comandou uma legião de entidades malignas, tendo todos sido lançados no inferno, tornando-se espíritos das trevas”.

Nossa proposta foi analisar como Paulo Coelho utiliza seu estilo lexical para construir as imagens de *Deus* e *demônio*. Para isto analisamos cada expressão em cotextos anteriores e posteriores às palavras analisadas. A decisão de utilizar palavras anteriores se deve ao fato de elas preparar o leitor para a palavra principal (*Deus* ou *demônio*), enquanto que as posteriores criam a imagem de Deus e demônio de acordo com suas ações. Destacamos que a expressão *Deus* aparece em 920 linhas de concordância e demônio em 112.

Figura 9: *Concord com as 20 primeiras ocorrências da palavra Deus no conjunto de obras de Paulo Coelho*

N	Concordance	Set	Tag	Word #	Sent	Sent Para	Para	Hear	Hear Sect	File	Date	%
1	não conseguia se lembrar — rezava e pedia a Deus que lhe ajudasse. Em nenhum			4.865	241	86%	1	10%	0	11%	O demônio e a : 2021/abr/01 00:1	11%
2	, calçada, tudo está em seu lugar, graças a Deus. Pegam em algum lugar do salão uma			62.357	3.6	100	0	56%	0	56%	O vencedor está 2021/abr/01 00:1	56%
3	duas maneiras de demonstrar meu amor a Deus: a primeira é adorá-Lo dia e noite, na			58.675	3.2	32%	0	98%	0	98%	Hippie.txt 2021/abr/01 00:1	98%
4	de disciplina, rigor, um lugar para servir a Deus de modo que pudesse estar mais perto			58.141	3.2	64%	0	97%	0	97%	Hippie.txt 2021/abr/01 00:1	97%
5	. Trabalhavam com honestidade, temiam a Deus, e procuravam ajudar o próximo. Faziam			1.528	129	58%	12	83%	0	3%	Brida.txt 2021/abr/01 00:1	3%
6	é apenas um dos muitos que levam a Deus. Jesus disse certa vez: “A casa de			50.106	3.3	100	1	38%	0	83%	O diário de um i 2021/abr/01 00:1	83%
7	ganham. “Li muito. Frequentei a igreja. Ternei a Deus, respeitei seus mandamentos. Era um			14.249	703	57%	1	32%	0	33%	O demônio e a : 2021/abr/01 00:1	33%
8	. Tudo isso por que razão? Porque temia a Deus e sabia que o ca- minho que percorreria			51.053	2.9	19%	0	46%	0	46%	O vencedor está 2021/abr/01 00:1	46%
9	mãos, para cuidar dos pobres e louvar a Deus. “Depois que caí em si e percebi o			10.894	645	100	0	25%	0	18%	O diário de um i 2021/abr/01 00:1	18%
10	, fez uma longa oração e entregou a guarda a Deus. “No dia seguinte, quando acordaram,			10.793	1.0	100	25	26%	0	27%	As valquírias.txt 2021/abr/01 00:1	27%
11	, e aprender com eles. Pediu inspiração a Deus, e começou a rezar. Duas horas depois			44.734	2.6	56%	0	40%	0	40%	O vencedor está 2021/abr/01 00:1	40%
12	permitido em nome do amor. Mas, graças a Deus, o seu momento mais difícil já ficara			79.297	4.6	36%	0	72%	0	72%	O vencedor está 2021/abr/01 00:1	72%
13	. Porque sabe que não está se igualando a Deus. Porque tem certeza de que a sua			66.838	3.9	100	0	60%	0	60%	O vencedor está 2021/abr/01 00:1	60%
14	e passa a chamá-lo Pedro. Moisés pergunta a Deus o seu nome: “Eu sou”, é a resposta.			13.945	801	38%	0	13%	0	13%	O vencedor está 2021/abr/01 00:1	12%
15	reagir é procurar servir aos outros, mostrar a Deus que sou uma pessoa boa, que mereço			73.785	4.4	76%	0	97%	0	97%	O zahir.txt 2021/abr/01 00:1	97%
16	a minha força e toda a minha fé: vou pedir a Deus que não permita passar o resto de			55.245	3.2	65%	0	72%	0	72%	O zahir.txt 2021/abr/01 00:1	72%
17	festa, jantaram com eles, dando graças a Deus por Dublin ter sempre um restaurante			16.840	1.3	70%	177	86%	0	33%	Brida.txt 2021/abr/01 00:1	34%
18	família, gostaria de ter filhos, honrar e temer a Deus. Os filhos, entretanto, não vieram. Olívia			4.215	214	100	0	4%	0	4%	O vencedor está 2021/abr/01 00:1	4%
19	presente no pensamento; e disso, graças a Deus, Maria já conseguira se livrar. Mesmo			39.232	2.1	82%	0	63%	0	63%	Onze minutos.t 2021/abr/01 00:1	63%
20	eu chegara a pensar em pedir perdão a Deus e romper com minha promessa. Mas ali			59.042	3.2	72%	0	95%	0	95%	Onze minutos.t 2021/abr/01 00:1	95%

Fonte: Elaboração própria.

²⁴ Disponível em <https://michaelis.uol.com.br/>. Acesso em 27 nov 2021.

Figura 10: Concord com as 20 primeiras ocorrências da palavra *demônio* no conjunto de obras de Paulo Coelho

N	Concordance	Set	Tag	Word #	Sent	Para	Para	Heai	Heai	Sect	Sect	File	Date	%
1	e o que estou fazendo aqui. Perdoe você e o demônio que tocava meu corpo quando eu	36.227	2.6:	29%	0	57%	0	57%	0	57%	0	O aleph.txt	2021/abr/01 00:1	57%
2	homens reúnem-se em um canto, eu escuto o demônio gritando cada vez mais alto nos	44.439	3.2:	39%	0	70%	0	70%	0	70%	0	O aleph.txt	2021/abr/01 00:1	70%
3	de um grande passeio no mundo', responde o demônio. "Então, viste o meu servo Jó? Viste	28.061	1.5:	100	1	65%	0	66%	0	66%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	66%
4	Paulo. - Deus está nas palavras, e o demônio também. Vahalla riu. - Parece que	17.696	1.6:	90%	25	44%	0	45%	0	45%	0	As valquírias.txt	2021/abr/01 00:1	45%
5	Não quero vê-la sofrendo. - Nunca invoquei o demônio. Você me conhece e também sabe	42.638	3.1:	100	0	67%	0	67%	0	67%	0	O aleph.txt	2021/abr/01 00:1	67%
6	de segunda-feira, dia em que invocavam o demônio. As cerimônias eram conduzidas	56.997	3.6:	100	0	86%	0	86%	0	86%	0	A bruxa de Port	2021/abr/01 00:1	85%
7	coisa com os outros. Isso é justiça. O demônio assustou-se, mas resolveu ficar	19.913	994	18%	1	46%	0	47%	0	47%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	47%
8	o silêncio. - Petrus, acho que o cigano era o demônio. - Sim, ele era o demônio - e	4.766	262	100	0	11%	0	8%	0	8%	0	O diário de um :	2021/abr/01 00:1	8%
9	por fazer. Mas vinha com sua companhia: o demônio. "Meu marido tem razão", disse para	117	5	100	0	15%	0	0%	0	0%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	0%
10	ciência que se opõe aos mistérios da fé. O demônio sabe que não pode trabalhar sozinho	41.684	3.0:	14%	0	65%	0	65%	0	65%	0	O aleph.txt	2021/abr/01 00:1	66%
11	que o cigano era o demônio. - Sim, ele era o demônio - e quando confirmou isto, senti um	4.772	263	37%	0	11%	0	8%	0	8%	0	O diário de um :	2021/abr/01 00:1	8%
12	se transformou em um adulto responsável? O demônio foi perdoado pela sua arrogância, e	103.612	6.1:	20%	0	93%	0	93%	0	93%	0	O vencedor está	2021/abr/01 00:1	94%
13	me adora, e faz todos os seus sacrifícios? "O demônio ri e argumenta: 'afinal de contas, Jó	28.080	1.5:	14%	1	65%	0	66%	0	66%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	66%
14	que isso o deixe mais tranquilo", escutara o demônio dizer. "Todos estão aterrorizados;	19.072	959	92%	1	44%	0	45%	0	45%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	45%
15	está entendendo, e está dizendo que sim." O demônio do estrangeiro estava sentindo-se	27.003	1.4:	12%	1	62%	0	63%	0	63%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	63%
16	: tinha escalado uma cachoeira, e derrotado o demônio do Caminho. Agora faltava apenas a	44.568	2.9:	92%	1	5%	0	74%	0	74%	0	O diário de um :	2021/abr/01 00:1	74%
17	o Bom Combate. "Quando não se conhece o demônio pessoal, ele costuma manifestar-se	15.346	931	47%	0	35%	0	25%	0	25%	0	O diário de um :	2021/abr/01 00:1	25%
18	. - Afaste-se em nome de Jesus! Gritei para o demônio, mas sem querer empurrei a	43.478	3.1:	31%	0	68%	0	68%	0	68%	0	O aleph.txt	2021/abr/01 00:1	68%
19	silêncio. "Não estou gostando nada", pensou o demônio do estrangeiro, que já começava a	27.443	1.4:	27%	1	63%	0	64%	0	64%	0	O demônio e a : 2021/abr/01 00:1	2021/abr/01 00:1	64%
20	fazendo aquilo para vingar-se do mundo. O demônio insiste, mas ele precisa mostrar-se	64.053	3.7:	19%	0	58%	0	58%	0	58%	0	O vencedor está	2021/abr/01 00:1	58%

Fonte: Elaboração própria.

3.1.4 Usando o Índice Flesch

Nesta pesquisa utilizamos cálculos de índices Flesch de cada obra de Paulo Coelho e demais autores, fornecidos pela página NILC-Metrix, da USP²⁵. Com os valores apresentados, tivemos condições de estipular as dificuldades de leitura das obras, ou seja, a CT.

Finatto (2011) apresenta uma escala com classificações quanto à CT, entre 0 e 100, porém mensurada com as devidas adaptações para o sistema escolar brasileiro, feitas pela equipe PorSimples²⁶. Os textos com números mais próximos de 100 são mais fáceis de compreender, enquanto os mais próximos de zero são os mais difíceis.

Quadro 9 – Níveis de inteligibilidade de acordo com o índice Flesch

COMPLEXIDADE TEXTUAL	ÍNDICE FLESH	PÚBLICO LEITOR
Muito fácil	Entre 75 e 100	Textos adequados para leitores com nível de escolaridade até a quarta série ²⁷ do ensino fundamental.

²⁵ Disponível em <http://fw.nilc.icmc.usp.br:23380/metrixdoc#flesch>. Acesso em 4 jul. 2021.

²⁶ O projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital) propõe o desenvolvimento de uma tecnologia para facilitar o acesso à informação dos analfabetos funcionais (AF) e, potencialmente, de pessoas com outras deficiências cognitivas, como afasia e dislexia. Disponível em <http://nilc.icmc.usp.br/simplifica/sobre.php>. Acesso em 4 jul. 2021.

²⁷ Segundo o Ministério da Educação (2009), a atual estruturação do ensino fundamental no Brasil foi iniciada em 2005. Desde então, a equivalência entre o ensino fundamental de oito e nove anos de duração passou a substituir a nomenclatura série escolar por ano escolar. Sendo assim, a quarta série foi renomeada como quinto ano e a oitava série foi renomeada como nono ano. Informações disponíveis em <http://portal.mec.gov.br/ensino-fundamental-de-nove-anos> e http://portal.mec.gov.br/dmdocuments/passos_a_passo_versao_atual_16_setembro.pdf. Acessos em 4 jul.2021.

Fácil	Entre 50 e 75	Textos adequados a alunos com escolaridade até a oitava série do ensino fundamental.
Difícil	Entre 25 e 50	Textos adequados para alunos cursando o ensino médio ou universitário.
Muito difícil	Entre 0 e 25	Textos adequados apenas para áreas acadêmicas específicas.

Fonte: Finatto, 2011

Inicialmente desejávamos utilizar um índice Flesch total de cada obra constante na pesquisa, mas a página do NILC-Metrix aceita no máximo duas mil palavras para cada cálculo. Existia a opção de dividir cada obra a cada duas mil palavras, o que demandaria muito tempo, por exemplo, *O alquimista*, de Paulo Coelho, resultaria em 19 divisões e cálculos. Optamos por utilizar as primeiras duas mil palavras de cada obra, duas mil localizadas no meio e últimas duas mil. Os trechos do meio de cada obra foram obtidos dividindo o total de páginas por dois e coletando aquelas que se encontravam no trecho mediano. Em seguida, somamos os três valores apresentados para cada cálculo e dividimos por três para obtermos uma média do índice Flesch.

4 ANÁLISES E RESULTADOS

As análises apresentadas nesta seção foram baseadas em resultados apresentados pelo WST e índice Flesch, constantes nos apêndices ao final desta pesquisa.

4.1 Comparativos entre *O alquimista*, de Paulo Coelho, e obras de demais autores

Nesta seção analisamos aspectos e resultados que colaboram para constituir o estilo lexical de Paulo Coelho e para isto promovemos comparações entre *O alquimista* e obras de outros autores. Na tabela a seguir constam dados das obras analisadas, sendo eles *tokens*, *types*, STTR e índices Flesch.

Tabela 1 – Tabela comparativa entre *O alquimista*, de Paulo Coelho, e obras de outros autores, em ordem alfabética pelos nomes dos autores

AUTOR	OBRA	ANO	TIPO DE TEXTO	TOKENS	TYPES	STTR	ÍNDICE FLESCH ²⁸
Aluísio Azevedo	<i>O cortiço</i>	1890	Romance	81.735	11.178	50,69	55,95373
Álvares de Azevedo	<i>Lira dos vinte anos</i>	1853	Poema	31.893	5.480	47,86	80,32387
Antônio Vieira	<i>Sermão de Santo Antônio aos peixes</i>	1682	Sermão	12.104	2.717	42,79	67,1863
Castro Alves	<i>Espumas flutuantes</i>	1870	Poema	18.582	4.628	52,72	78,21337
Chico Buarque	<i>Budapeste</i>	2003	Romance	40.445	8.079	51,24	67,21297
Cruz e Sousa	<i>Broquéis</i>	1893	Poema	5.369	2.068	56,66	57,16695
Euclides da Cunha	<i>Os sertões</i>	1902	Romance	154.670	22.092	56,58	30,31280
Graciliano Ramos	<i>Vidas secas</i>	1938	Romance	25.030	5.572	53,76	50,27701
Gregório de Matos	<i>Seleção de obras poéticas</i> (antologia)	Antologia	Poema	5.417	1.823	50,16	80,05619
João Guimarães Rosa	<i>Grande sertão: veredas</i>	1956	Romance	189.162	19.973	50,09	78,75608
José de Alencar	<i>Iracema</i>	1865	Romance	22.466	3.877	45,01	69,67011
José de Anchieta	<i>Auto representado na Festa de São Lourenço</i>	1583	Peça teatral	6.869	2.200	50,53	76,56316
Machado de Assis	<i>Dom Casmurro</i>	1899	Romance	66.807	8.683	47,36	67,15932
Mário de Andrade	<i>Macunaíma</i>	1928	Romance	43.185	8.085	47,65	67,26008
Olavo Bilac	<i>Alma inquieta</i> (antologia)	Antologia	Poema	9.514	2.495	47,81	80,94455
Paulo Coelho	<i>O alquimista</i>	1988	Romance	35.228	4.285	42,79	63,18681
Pero Vaz de Caminha	<i>A carta</i>	1500	Carta	8.033	1.694	42,61	70,26244
Tomás Antônio Gonzaga	<i>Marília de Dirceu</i>	1792	Poema	21.216	4.298	49,98	84,16514

Fonte: Elaboração própria

²⁸ Obtido com a soma de três índices Flesch de trechos da obra (início, meio e fim), cujo resultado foi dividido por três.

Pela coluna *tokens* constatamos que a obra *Broquéis*, conta com a menor quantidade, 5.369 ao todo, seguida de *Seleção de obras poéticas* (5.417); *Auto representado na Festa de São Lourenço* (6.869); *A carta*, (8.033); *Alma inquieta* (9.514); *Sermão de Santo Antônio aos peixes* (12.104); *Espumas flutuantes* (18.582); *Marília de Dirceu* (21.216); *Iracema* (22.466); *Vidas secas* (25.030), *Lira dos vinte anos* (31.893); *O alquimista* (35.228); *Budapeste* (40.445), *Macunaíma* (43.185), *Dom Casmurro* (66.807); *O cortiço* (81.735), *Os sertões* (154.670) e *Grande sertão: veredas* (189.162).

Os textos do tipo poema, carta, sermão e peça teatral contam com o menor número de *tokens*, a única exceção é a coletânea de poemas *Lira dos vinte anos*, classificada após os romances *Iracema* e *Vidas secas*. A maioria das obras com maior quantidade de *tokens* é constituída por romances, sendo a única exceção os poemas de *Lira dos vinte anos*.

Quanto ao romance *O alquimista*, ele possui 35.228 *tokens*, uma quantidade maior que os romances *Iracema* e *Vidas secas* e que todos os demais textos escritos em outros gêneros. Entretanto, *O alquimista* é menor em número de *tokens* em relação aos romances *Budapeste*, *Macunaíma*, *Dom Casmurro*, *O cortiço*, *Os sertões* e *Grande sertão: veredas*.

A coluna *types* apresenta, em ordem crescente, *A carta* (1.694), *Seleção de obras poéticas* (1.823); *Broquéis* (2.068); *Auto representado na Festa de São Lourenço* (2.200); *Alma inquieta* (2.495); *Sermão de Santo Antônio aos peixes* (2.717); *Iracema* (3.877); *O alquimista* (4.285); *Marília de Dirceu* (4.298); *Espumas flutuantes* (4.628); *Lira dos vinte anos* (5.480); *Vidas secas* (5.572), *Budapeste* (8.079), *Macunaíma* (8.085), *Dom Casmurro* (8.683); *O cortiço* (11.178), *Grande sertão: veredas* (19.973) e *Os sertões* (22.092).

Esta coluna revela uma maior diversidade entre os gêneros dos textos. Entre os romances, sete constam como as obras com o maior número de *types*, sendo eles, *Vidas secas*, *Budapeste*, *Macunaíma*, *Dom Casmurro*, *O cortiço*, *Grande sertão: veredas* e *Os sertões*. Entre os romances, *O alquimista* tem mais *types* apenas que *Iracema*, porém menos que os poemas de *Marília de Dirceu*, *Espumas flutuantes* e *Lira dos vinte anos*.

A coluna STTR apresenta, em ordem crescente de densidade lexical de cada obra, *A carta* (42,61), *Sermão de Santo Antônio aos peixes* (42,79), *O alquimista*, (42,79), *Iracema* (45,01), *Dom Casmurro* (47,36), *Macunaíma* (47,65), *Alma inquieta* (47,81), *Lira dos vinte anos* (47,86), *Marília de Dirceu* (49,98), *Grande sertão: veredas* (50,09), *Seleção de obras poéticas* (50,16), *Auto representado na Festa de São Lourenço* (50,53), *O cortiço* (50,69),

Budapeste (51,24), *Espumas flutuantes* (52,72), *Vidas secas* (53,76), *Os sertões* (56,58) e *Broquéis* (56,66).

Quando se observa STTR, percebe-se que *O alquimista* conta com um nível de densidade lexical inferior a quase todas as obras. Em uma lista em ordem crescente, fica na segunda classificação, em empate com o *Sermão de Santo Antônio aos Peixes* e maior apenas que *A Carta*. Percebemos uma mescla quanto aos gêneros dos textos quando classificamos todas as obras. Diante disso, a densidade lexical de Paulo Coelho, em *O alquimista*, revelou-se inferior se comparada às dos demais autores.

Em outra análise, pela coluna do índice Flesch tivemos conhecimento da CT de cada obra diante da facilidade ou dificuldade de compreensão de um texto pelo leitor, graças a itens como coesão e coerência.

Apresentamos a seguir uma tabela comparativa, em ordem crescente de índice Flesch, entre *O alquimista* e obras de demais autores. Como esclarecemos anteriormente, quanto maior o resultado do índice Flesch, menor a CT e mais fácil é a compreensão. Do contrário, quanto menor o resultado do índice Flesch, maior é a CT e mais difícil é a compreensão.

Tabela 2 – Tabela comparativa entre *O alquimista*, de Paulo Coelho, e as obras dos demais autores, apresentada em ordem crescente de índice Flesch

AUTOR	OBRA	TIPO DE TEXTO	ÍNDICE FLESCH ²⁹
Difícil – Índice Flesch entre 25 e 50			
Euclides da Cunha	<i>Os sertões</i>	Romance	30,31280
Fácil – Índice Flesch entre 50 e 75			
Graciliano Ramos	<i>Vidas secas</i>	Romance	50,27701
Aluísio Azevedo	<i>O cortiço</i>	Romance	55,95373
Cruz e Sousa	<i>Broquéis</i>	Poema	57,16695
Paulo Coelho	<i>O alquimista</i>	Romance	63,18681
Machado de Assis	<i>Dom Casmurro</i>	Romance	67,15932
Antônio Vieira	<i>Sermão de Santo Antônio aos peixes</i>	Sermão	67,1863
Chico Buarque	<i>Budapeste</i>	Romance	67,21297
Mário de Andrade	<i>Macunaíma</i>	Romance	67,26008
José de Alencar	<i>Iracema</i>	Romance	69,67011
Pero Vaz de Caminha	<i>A carta</i>	Carta	70,26244
Muito fácil – Entre 75 e 100			
José de Anchieta	<i>Auto representado na Festa de São Lourenço</i>	Peça teatral	76,56316
Castro Alves	<i>Espumas flutuantes</i>	Poema	78,21337
João Guimarães	<i>Grande sertão: veredas</i>	Romance	78,75608

²⁹ Obtido com a soma de três índices Flesch de trechos da obra (início, meio e fim), cujo resultado foi dividido por três.

Rosa			
Gregório de Matos	<i>Seleção de obras poéticas (antologia)</i>	Poema	80,05619
Álvares de Azevedo	<i>Lira dos vinte anos</i>	Poema	80,32387
Olavo Bilac	<i>Alma inquieta (antologia)</i>	Poema	80,94455
Tomás Antônio Gonzaga	<i>Marília de Dirceu</i>	Poema	84,16514

Fonte: Elaboração própria

De acordo com o índice Flesch e a CT, a única obra considerada de difícil compreensão é o romance *Os sertões* (30,31280).

Na classificação de fácil inteligibilidade constam os romances *Vidas secas* (50,27701), *O cortiço* (55,95373), *O alquimista* (63,18681), *Dom Casmurro* (67,15932), *Budapeste* (67,21297), *Macunaíma* (67,26008) e *Iracema* (69,67011); além dos poemas de *Broquéis* (57,16695), da peça teatral *Sermão de Santo Antônio aos peixes* (67,1863) e de *A carta* (70,26244),

As obras consideradas com compreensão muito fácil são a peça teatral *Auto representado na Festa de São Lourenço* (76,56316), o poema *Espumas flutuantes* (78,21337), o romance *Grande sertão: veredas* (78,75608) e os poemas *Seleção de obras poéticas* (80,05619), *Lira dos vinte anos* (80,32387), *Alma inquieta* (80,94455) e *Marília de Dirceu* (84,16514).

A obra *O alquimista*, um dos nossos objetos de estudo, é classificada como de fácil compreensão. Em comparação com outros romances, é mais fácil que *Os sertões*, *Vidas secas* e *O cortiço*, porém mais difícil que *Dom Casmurro*, *Budapeste*, *Macunaíma*, *Iracema* e *Grande sertão: veredas*. Quanto aos poemas, é mais fácil de ser compreendida que *Broquéis* e mais difícil que *Espumas flutuantes*, *Seleção de obras poéticas*, *Lira dos vinte anos*, *Alma inquieta* e *Marília de Dirceu*. Em comparação com outros tipos de textos, é mais difícil que *A carta*, *Sermão de Santo Antônio aos Peixes* e *Auto representado na Festa de São Lourenço*.

De acordo com o índice Flesch, *O alquimista* apresenta um nível fácil de compreensão pelos leitores. Destacamos ainda que a obra encontra-se em nível semelhante aos demais romances *Vidas secas*, *O cortiço*, *Dom Casmurro*, *Budapeste*, *Macunaíma* e *Iracema*. O que a credencia como uma obra detentora de uma compreensão e CT condizentes com algumas outras obras famosas da literatura brasileira.

4.2 Comparativos entre as obras de Paulo Coelho

A tabela a seguir apresenta uma comparação entre as obras de Paulo Coelho, listadas em ordem alfabética e com escalas de *tokens*, *types*, STTR e índice Flesch.

Tabela 3 – Tabela comparativa entre as obras de Paulo Coelho em ordem alfabética de título

OBRA	ANO	TOKENS	TYPES	STTR	ÍNDICE FLESCH ³⁰
<i>A bruxa de Portobello</i>	2006	65.217	8.272	49,17	55,34700
<i>A espiã</i>	2016	29.146	5.622	50,35	54,36032
<i>Adultério</i>	2014	59.586	8.158	49,63	65,08749
<i>As valquírias</i>	1992	38.068	5.234	46,41	62,12658
<i>Brida</i>	1990	49.922	5.870	45,27	67,08166
<i>Hippie</i>	2018	58.952	8.075	49,27	49,55452
<i>Na margem do rio Piedra eu sentei e chorei</i>	1994	35.359	4.881	47,07	72,50551
<i>O Aleph</i>	2010	62.255	7.799	48,23	66,77339
<i>O alquimista</i>	1988	35.228	4.285	42,79	63,18681
<i>O demônio e a senhorita Prym</i>	2000	41.746	6.074	48,28	60,67665
<i>O diário de um mago</i>	1987	59.395	7.182	46,02	61,95308
<i>O monte Cinco</i>	1996	40.047	5.811	47,88	64,31547
<i>O vencedor está só</i>	2008	109.002	12.142	51,39	53,98594
<i>O zahir</i>	2005	74.428	8.538	47,94	63,77542
<i>Onze minutos</i>	2003	61.153	7.291	47,56	56,27822
<i>Veronika decide morrer</i>	1998	46.534	6.639	48,76	51,38198

Fonte: Elaboração própria

A coluna *tokens* apresenta a obra *A espiã* com o menor total de *tokens* (29.146), seguida por *O alquimista* (35.228), *Na margem do rio Piedra eu sentei e chorei* (35.359), *As valquírias* (38.068), *O monte Cinco* (40.047), *O demônio e a senhorita Prym* (41.746), *Veronika decide morrer* (46.534), *Brida* (49.922), *Hippie* (58.952), *O diário de um mago* (59.395), *Adultério* (59.586), *Onze minutos* (61.153), *O Aleph* (62.255), *A bruxa de Portobello* (65.217), *O zahir* (74.428) e *O vencedor está só* (109.002).

Com exceção de *A espiã* e *O vencedor está só*, as demais obras possuem entre 35.228 e 74.428 *tokens*. A obra mais antiga da lista e do ano de 1987, *O diário de um mago*, apresenta 59.395 *tokens*, enquanto *Hippie*, a mais atual e do ano de 2018, conta com 58.952, revelando uma pequena variabilidade na quantidade de *tokens* quando observamos a ordem cronológica dos lançamentos.

³⁰ Obtido com a soma de três índices Flesch de trechos da obra (início, meio e fim), cujo resultado foi dividido por três.

Pela coluna *types* temos como resultados em ordem crescente, *O alquimista* (4.285), *Na margem do rio Piedra eu sentei e chorei* (4.881), *As valquírias* (5.234), *A espiã* (5.622), *O monte Cinco* (5.811), *Brida* (5.870), *O demônio e a senhorita Prym* (6.074), *Veronika decide morrer* (6.639), *O diário de um mago* (7.182), *Onze minutos* (7.291), *O Aleph* (7.799), *Hippie* (8.075), *Adultério* (8.158), *A bruxa de Portobello* (8.272), *O zahir* (8.538) e *O vencedor está só* (12.142).

Considerando as datas dos lançamentos das obras, a variação não segue uma ordem cronológica, pois *O diário de um mago*, de 1987, conta com 7.182 *types* e *Hippie*, de 2018, apresenta 8.075. Sendo assim, entre a obra mais antiga e a mais recente existe uma variação de 12,43%. Com exceção de *O vencedor está com só*, que apresenta 12.142 *types*, as demais obras contam com um total de *types* entre 4.285 e 8.538, revelando uma variação da ordem de 99,25%. Incluindo *O vencedor está só*, a variação lexical é de 183,36%.

Pela STTR apresentada, *O alquimista*, de 1988, é a obra com menor densidade lexical (42,79), seguida por *Brida* (45,27), *O diário de um mago* (46,02), *As valquírias* (46,41), *Na margem do rio Piedra eu sentei e chorei* (47,07), *Onze minutos* (47,56), *O monte Cinco* (47,88), *O zahir* (47,94), *O Aleph* (48,23), *O demônio e a senhorita Prym* (48,28), *Veronika decide morrer* (48,76), *A bruxa de Portobello* (49,17), *Hippie* (49,27), *Adultério* (49,63), *A espiã* (50,35) e *O vencedor está só* (51,39).

Ao observarmos as datas de lançamento das obras, a mais antiga, *O diário de um mago*, de 1987, atinge uma densidade lexical de 46,02, enquanto a mais recente, *Hippie*, de 2018, conta com 49,27, um aumento de 7,06%. Quanto à variação entre *O alquimista* que obtém a menor densidade lexical (42,79) e *O vencedor está só*, com a maior densidade lexical (51,39), o aumento constatado é de 20,09%.

A seguir apresentamos uma tabela comparativa, em ordem crescente de índice Flesch, entre as obras de Paulo Coelho.

Tabela 4 – Tabela comparativa entre as obras de Paulo Coelho, apresentada em ordem crescente de índice Flesch

OBRA	ANO	ÍNDICE FLESCH ³¹
Difícil – Índice Flesch entre 25 e 50		
<i>Hippie</i>	2018	49,55452
Fácil – Índice Flesch entre 50 e 75		
<i>Veronika decide morrer</i>	1998	51,38198
<i>O vencedor está só</i>	2008	53,98594
<i>A espiã</i>	2016	54,36032

³¹ Obtido com a soma de três índices Flesch de trechos da obra (início, meio e fim), cujo resultado foi dividido por três.

<i>A bruxa de Portobello</i>	2006	55,34700
<i>Onze minutos</i>	2003	56,27822
<i>O demônio e a senhorita Prym</i>	2000	60,67665
<i>O diário de um mago</i>	1987	61,95308
<i>As valquírias</i>	1992	62,12658
<i>O alquimista</i>	1988	63,18681
<i>O zahir</i>	2005	63,77542
<i>O monte Cinco</i>	1996	64,31547
<i>Adultério</i>	2014	65,08749
<i>O Aleph</i>	2010	66,77339
<i>Brida</i>	1990	67,08166
<i>Na margem do rio Piedra eu sentei e chorei</i>	1994	72,50551

Fonte: Elaboração própria

Pelos índices Flesch apresentados, percebe-se que a única obra com compreensão considerada difícil é *Hippie* (49,55452).

Quanto às demais obras, são todas classificadas com compreensão fácil, sendo elas, *Veronika decide morrer* (51,38198), *O vencedor está só* (53,98594), *A espiã* (54,36032), *A bruxa de Portobello* (55,34700), *Onze minutos* (56,27822), *O demônio e a senhorita Prym* (60,67665), *O diário de um mago* (61,95308), *As valquírias* (62,12658), *O alquimista* (63,18681), *O zahir* (63,77542), *O monte Cinco* (64,31547), *Adultério* (65,08749), *O Aleph* (66,77339), *Brida* (67,08166) e *Na margem do rio Piedra eu sentei e chorei* (72,50551)

Não constam obras com inteligibilidade consideradas muito fácil ou muito difícil.

Pelo índice Flesch conclui-se que a maioria das obras de Paulo Coelho conta com um texto de fácil compreensão, permitindo uma leitura acessível à grande parte dos leitores.

4.3 Análises das *hapax legomena*, palavras-chave e dos campos lexicais das obras de Paulo Coelho

Ao utilizarmos uma *Wordlist*, produzida pelo WST, tivemos conhecimento da quantidade das *hapax legomena* presentes nos 16 romances de Paulo Coelho. O resultado apresentado é que em um total de 32.793 palavras diferentes utilizadas, 13.505, ou seja, 41,18%, são *hapax legomena*. Conforme Grama (2016), isto é importante para mensurar a densidade lexical do autor. Deduzimos, então, que Paulo Coelho conta com uma ampla densidade lexical ao se observar o uso de *hapax legomena*, pois quase metade do léxico utilizado por ele pertence a este grupo.

Algo que merece ser mencionado quanto ao uso de *hapax legomena* por Paulo Coelho é o uso do verbo vingar. Segundo o Dicionário Michaelis³², vingar provém de vingança, um ato praticado contra outra pessoa em castigo diante de uma ofensa recebida, causando dano semelhante. Constatou-se ser uma ação negativa deferida contra outra pessoa. Paulo Coelho se utiliza das expressões *vingá-los*, *vingando-se*, *vingar-me*, *vingou* e *vinguei* apenas uma vez no conjunto de todas as suas obras.

A partir de outra lista, porém de *Keywords*, também fornecida pelo WST e produzida com os romances de Paulo Coelho, obtivemos as palavras-chave usadas pelo autor. Totalizando 187 palavras-chave (substantivos e adjetivos), categorizamos as mesmas em grupos de acordo com definições apresentadas nas versões digitais dos dicionários *Aulete Digital* (2007), *Dicio* (2009) e *Michaelis* (2015). Em algumas situações recorremos a outras fontes, como a *Bíblia Sagrada* (1990) e o endereço eletrônico *Significados*³³. Após as consultas, as palavras foram agrupadas em campos lexicais de acordo com suas áreas de concentração e hiperônimos, sendo que algumas são intercambiáveis e constam em mais de um campo lexical.

Na figura abaixo constam as 20 primeiras palavras-chave de um recorte constituído por substantivos e adjetivos, do conjunto das obras de Paulo Coelho.

Figura 11 – *Keyword*, com *stoplist*, com as 20 primeiras palavras-chave (substantivos e adjetivos) de todas as obras de Paulo Coelho, cujas chavidades constam em ordem crescentes na coluna *Keyness*



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas	Set
1	AMOR	1.340	0,15	16	50.131		5.627,95	0,0000000000		
2	RAPAZ	670	0,08	15	9.971		3.984,83	0,0000000000		
3	MULHER	1.469	0,17	16	131.179	0,02	3.826,14	0,0000000000		
4	HOMEM	1.536	0,18	16	147.561	0,02	3.803,34	0,0000000000		
5	OLHOS	908	0,10	16	39.178		3.570,24	0,0000000000		
6	COISA	1.305	0,15	16	117.630	0,02	3.377,23	0,0000000000		
7	COISAS	1.106	0,13	16	85.403	0,01	3.166,63	0,0000000000		
8	VIDA	2.104	0,24	16	372.470	0,05	3.072,53	0,0000000000		
9	WICCA	229	0,03	1	0		3.063,10	0,0000000000		
10	MUNDO	1.902	0,22	16	309.279	0,04	3.031,31	0,0000000000		
11	MEDO	749	0,09	16	41.449		2.598,67	0,0000000000		
12	DEUS	920	0,11	16	77.069	0,01	2.499,99	0,0000000000		
13	DESERTO	373	0,04	13	4.364		2.388,33	0,0000000000		
14	ALMA	539	0,06	16	20.700		2.236,78	0,0000000000		
15	NOITE	935	0,11	16	95.516	0,01	2.214,54	0,0000000000		
16	CAMINHO	787	0,09	16	62.819		2.206,30	0,0000000000		
17	MOMENTO	1.321	0,15	16	208.888	0,03	2.163,60	0,0000000000		
18	VISCOS	145	0,02	1	0		1.939,50	0,0000000000		
19	PESSOAS	1.649	0,19	16	353.124	0,05	1.933,83	0,0000000000		
20	ANJO	309	0,04	16	4.176		1.894,17	0,0000000000		

Fonte: Tela do WST

³² Disponível em <<https://michaelis.uol.com.br/>>. Acesso em 27 fev. 2022.

³³ Disponível em <<https://www.significados.com.br/>>. Acesso em 18 jul. 2021.

A seguir, apresentamos alguns campos lexicais e destacamos após as palavras e entre os parênteses, respectivamente, a quantidade de textos nos quais as palavras ocorreram, a frequência das palavras e a chavicidade apresentada na coluna *Keyness*.

4.3.1 Campo lexical Religiosidades e Crenças

A identificação deste campo lexical teve como base a utilização de palavras que remetem os leitores a diversas religiões e crenças.

No primeiro subcampo lexical constam palavras que dizem respeito às divindades, sendo elas *Deus* (16 / 920 / 2.499,99), *deusa* (8 / 63 / 325,19), *deuses* (14 / 140 / 610,33), *anjo* (16 / 309 / 4,176), *anjos* (16 / 309 / 1.894,17). Mencionamos ainda a expressão *demônio* (11 / 112 / 2.467) pois é definida como um anjo que se rebelou contra Deus.

Locais relacionados às religiosidades formam o segundo subcampo lexical, sendo eles, *céu* (16 / 267 / 927,82), *céus* (16 / 100 / 502,11) e *paraíso* (15 / 117 / 310,07). Sendo *paraíso* abordado no *Antigo Testamento (Bíblia Sagrada, 1990)* como o jardim do Éden, criado por Deus para abrigar Adão e Eva; e no *Novo Testamento (Bíblia Sagrada, 1990)* como o lugar onde permanecem as almas dos homens bem-aventurados após a morte. Destacamos que *paraíso* é considerado sinônimo de *céu*.

Atribuições e profissões religiosas formam o terceiro subcampo lexical, sendo elas *padre* (11 / 223) e *sacerdote* (7 / 114 / 614,18).

Algumas palavras pertencem a culturas e locais específicos formando o quarto subcampo lexical. É o caso de *Aleph* (1 / 51 / 340,51), interpretado pelos adeptos das doutrinas cabalísticas como um símbolo místico e espiritual, representante de Deus como “o começo de tudo”. A expressão *Wicca* (1 / 229 / 3.063,1), que diz respeito a uma religião moderna baseada em rituais pagãos antigos, também chamada de bruxaria. A palavra *Zahir* (1 / 50 / 451,35) se refere a um conceito ligado à filosofia islâmica, que diz respeito aos comportamentos e atitudes exteriores dos indivíduos. A palavra intercambiável *ágape* (1 / 55 / 495,14) pertencente também ao campo lexical Sentimentos e é citada na *Bíblia Sagrada* (1990) referindo-se ao amor de Deus para com os homens e como uma refeição compartilhada entre os antigos cristãos.

Algumas palavras de cunho religioso constituem o quinto subgrupo, sendo elas *alma* (16 / 539 / 2.236,7), *milagres* (12 / 81 / 338,31), *perdão* (14 / 91 / 329,19), *peregrinos* (7 / 56 / 338,44) e *sagrado* (14 / 100 / 307,66).

Observamos que as palavras *Deus e anjo* são citadas em todas as 16 obras, enquanto *demônio* aparece em 11. *Deus* também é a palavra mais citada neste campo lexical, em um total de 920 vezes. Quanto às palavras *céu, paraíso, alma, perdão e sagrado* são citadas em 14 ou mais obras, ou seja, são constantes. Isto reforça a ideia de um texto com palavras que remetem a pensamentos positivos.

Quanto às palavras-chave *Aleph, Wicca, Zahir e ágape* são citadas em apenas um livro por estarem ligadas ao tema da obra ou a algum personagem.

4.3.2 Campo lexical Geografia

Este campo lexical é composto por palavras que dizem respeito a locais específicos ou que fazem parte dos estudos da Geografia.

O primeiro subcampo lexical contém palavras que remetem a locais específicos, como países, regiões e povos. Tais palavras são *assírios*, que são pessoas pertencentes à antiga Assíria, na Ásia (1 / 72 / 963,06), *Baikal*, lago na Sibéria, na Ásia (2 / 24 / 321,02), *Lubljana*, capital da Eslovênia, na Europa (1 / 31 / 414,65), *Santiago*, região da Espanha, na Europa (12 / 182 / 562,04), *Nepal*, país na Ásia (3 / 51 / 300,52), *Sidon*, cidade do Líbano, na Ásia (1 / 33 / 441,40), *Viscos*, cidade da França, na Europa (1 / 146 / 1.939,5) e *Villette*, sanatório na Eslovênia, na Europa (1 / 75 / 1.003,1).

O segundo subcampo lexical diz respeito a locais genéricos, definidos como de domínio da Geografia, sendo eles *cidade* (16 / 802 / 415,06), *deserto* (13 / 373 / 2.388,3), *oásis* (3 / 59 / 371,06), *montanha* (16 / 113 / 458,94) e *montanhas* (15 / 219 / 1.158,7).

Neste campo lexical, os locais específicos são citados em poucas obras, no máximo três. A única exceção é Santiago, região da Espanha, na Europa, onde Paulo Coelho conta ter peregrinado. Isto agrega uma conotação intimista e autobiográfica de Paulo Coelho em suas obras. Interessante também observarmos que todos os locais encontram-se na Europa e na Ásia e não são frequentemente mencionados pelos brasileiros, como a Torre Eiffel, na França, e a Grande Muralha, na China.

4.3.3 Campo lexical Sentimentos

Este campo lexical apresenta palavras cujas definições estão ligadas a sentimentos, sendo elas: *amor* (16 / 1.340 / 5.627,9), *medo* (16 / 749 / 2.598,6), *contente* (16 / 231 /

1.485,1), *alegria* (16 / 220 / 621,77), *feliz* (16 / 187 / 430,02), *prazer* (15 / 189 / 346,62) e *ódio* (11 / 93 / 308,99). Uma expressão específica é *ágape* (1 / 55 / 495,14), palavra intercambiável já citada e pertencente também ao campo lexical Religiosidades e Crenças, cujo significado é amor de Deus para com os homens.

Observamos que quase todas as palavras são citadas em todas as obras ou na maioria delas. Identificamos que *amor* surge nas 16 obras e *ódio* em 11. A palavra *ágape* é a única mencionada apenas em uma obra, sendo ela *O diário de um mago*.

4.3.4 Campo lexical Anatomia

As palavras que compõem este campo lexical constam nos dicionários como da área de Anatomia, geralmente pertencentes ao corpo humano, sendo elas *olhos* (16 / 908 / 3.570,2), *cabeça* (16 / 559 / 1.335,3), *rosto* (16 / 297 / 1.176,7), *mão* (16 / 425 / 605,99), *mãos* (16 / 412 / 976,62), *cabelos* (16 / 160 / 607,94), *braços* (16 / 155 / 468,65), *pés* (16 / 160 / 426,46) e *coração* (16 / 566 / 1.850,1).

Neste campo lexical, todas as palavras são mencionadas em todas as obras.

4.3.5 Campo lexical Astronomia

Embora pequeno, este campo lexical apresenta palavras definidas como pertencentes à Astronomia e são *lua* (13 / 178 / 654,47), *sol* (16 / 448 / 1.469,2) e *universo* (16 / 217 / 370,75).

As palavras *sol* e *universo* são mencionadas em todas as 16 obras e *lua* em 13 obras.

4.4 Análises das palavras *Deus* e *demônio* e seus cotextos no conjunto das obras de Paulo Coelho

Ao utilizarmos duas listas geradas pelo programa *Concord*, do WST, e buscarmos as palavras *Deus* e *demônio*, podemos analisar como tais palavras são utilizadas em seus cotextos.

4.4.1 *Deus* e seus cotextos no conjunto das obras de Paulo Coelho

A expressão *Deus* é citada 920 vezes em todos os 16 romances literários de Paulo Coelho. Concluímos se tratar de uma construção da imagem de *Deus* como um ser no qual as pessoas têm grande crença. Citamos como exemplo trechos como “acredite que Deus é mãe”, “adorar a Deus”, “amor de Deus”, “aproximar-se de Deus”, “benção de Deus”, “caminho direto até Deus”, “comunhão com Deus”, “contato com Deus”, “encontro com Deus”, “glória de Deus” e “graças a Deus”. Também observamos o uso de *Deus* mencionado no feminino, como em “face feminina” de deus e “rosto feminino de Deus”. Desta forma percebemos que Paulo Coelho mostra Deus como um ser bom, compassivo e com grande credibilidade pelas pessoas.

Outras características de Deus o evidencia como atuante, onipotente, onisciente, onipresente e o início de tudo. Apresentamos a seguir alguns exemplos de trechos retirados de livros de Paulo Coelho que confirmam estas definições: “Deus aceitou teu combate” (*O Monte Cinco*, 1996), “Deus enviará seu exército de anjos” (*As valquíria*, 1992), “O poder de Deus está conosco a cada momento, Deus está acompanhando os seus passos – se não reza, embora Deus esteja sempre certo. Deus estava me escutando” (*Hippie*, 2018), “senti que Deus havia escutado minhas preces” (*A bruxa de Portobello*, 2006), “Deus não é vingança, Deus é amor. Pela glória de Deus” (*O diário de um mago*, 1987), “Deus nos criou para a felicidade” (*Na margem do rio Piedra eu sentei e chorei*, 1994) e “Deus ficou contente” (*O demônio e a Srta. Prym*, 2000).

4.4.2 *Demônio* e seus cotextos no conjunto das obras de Paulo Coelho

A expressão *demônio* é citada 112 vezes em 11 obras de um conjunto de 16 romances de Paulo Coelho. A construção da imagem de demônio é feita mostrando-o como um ser que precisa de alguém para se manifestar. Geralmente são pessoas fracas, pois demônio é mostrado como consequência de pensamentos e atos nocivos do homem consigo mesmo e ao próximo. Isto é confirmado por trechos retirados de livros de Paulo Coelho: “seu demônio pessoal” (*O diário de um mago*, 1987), “o demônio foi se apossando de sua alma” (*O demônio e a Srta. Prym*, 2000), “o demônio sabe que não pode trabalhar sozinho, precisa de suas feiticeiras e seus cientistas para seduzir e corromper o mundo” (*O Aleph*, 2010), “nunca invoquei o demônio” (*O Aleph*, 2010), “o demônio me tenta, me provoca” (*O Aleph*, 2010),

“foi enviado pelo demônio para desestabilizar algo que já estava frágil” (*Adultério*, 2014) e “o homem abriu a janela do seu quarto, e torceu para que o frio calasse por alguns momentos a voz do seu demônio” (*O demônio e a Srta. Pryn*, 2000).

Finalmente fazemos uso de três trechos maiores e que definem demônio:

- “Mas apesar disso, continua ele, o demônio permanecia neste mundo, usando todo tipo de artifício – inclusive a pretensa capacidade de ver o futuro – para enganar os fracos e afastar as pessoas da verdadeira fé.” (*O Zahir*, 2005).

- “O demônio é um profissional de primeiríssima qualidade, e assusta os fracos com sentimentos de medo, preocupações, impotência, desespero.” (*O vencedor está só*, 2008).

- “Na Tradição, o demônio é um espírito que não é bom nem mau, mas considerado guardião da maior parte dos segredos acessíveis ao homem, e com força e poder sobre as coisas materiais. Por ser o anjo caído, identifica-se com a raça humana e está sempre disposto a pactos e trocas de favores.” (*O diário de um mago*, 1987),

4.4.3 Deus e demônio no estilo lexical de Paulo Coelho

Após analisarmos as palavras *Deus* e *demônio* na obra de Paulo Coelho, concluímos que o autor se utiliza de seu estilo lexical para reforçar os conceitos comuns nas religiões cristãs, não apresentando algo novo.

Deus é bom na obra de Paulo Coelho e existe independente de alguém. Pelo contrário, as pessoas é que precisam de Deus para serem felizes, pois Deus é a própria felicidade. Deus é mostrado como um ser justo, milagroso, bondoso, amoroso e que está sempre disposto a realizar o bem. Devido a esta bondade latente de Deus, o homem quer estar próximo a ele, confiando e sendo grato.

A expressão *demônio* já é abordada como o algo ruim na obra de Paulo Coelho. Entretanto, ao contrário de Deus, o demônio precisa das pessoas para realizar suas proezas. Muitas vezes, o demônio não passa de um fruto de pensamentos fracos de alguém e só se manifesta nos fracos, enquanto Deus está em tudo. Desta forma, Paulo Coelho nos mostra que, ao encontrar alguém fragilizado, o demônio se manifesta. Caso contrário, diante de alguém firme em suas boas convicções e em sua fé, o demônio perde força e não encontra espaço para realizar suas obras.

5 CONSIDERAÇÕES FINAIS

Ao concluirmos nossa pesquisa, que resultou nesta “Proposta metodológica para análises literárias através da Linguística de Corpus: o caso de Paulo Coelho”, entendemos que promovemos uma conexão entre a Linguística e a Literatura. Desde o início defendemos que o diálogo entre as duas áreas é necessário e enriquecedor. Para atingirmos tal objetivo, nos valem de referenciais teóricos que consideramos como indispensáveis.

Inicialmente utilizamos estudos da área do Léxico, tendo Biderman (1978, 1987, 2001) como norte principal. Assim como enfatiza a autora, entendemos que o léxico é indispensável na perpetuação de um conhecimento que engloba desde a rotulação inicial de algo, passando pela transmissão do conhecimento e, finalmente, na fixação de tal conhecimento. Isto torna essencial a existência e a utilização de um léxico único, normatizado e formalizado entre um grupo de pessoas. Caso contrário, a linguagem se tornará incompreensível. Entretanto, como também já ressaltamos, cada pessoa faz uso do léxico com traços peculiares e de acordo com sua personalidade e ideologias, além de outros fatores que tornam único cada ser humano, assim como sua história de vida. Ao analisar obras literárias, esta proposta teve que obrigatoriamente observar o léxico empregado em tais produções.

Neste nosso percurso, estudar o estilo empregado nas obras literárias se tornou necessário e nos valem da definição de que estilo é escolha, conforme nos afirma diversos autores mencionados nesta pesquisa, principalmente Martins (2012), Morejón e Martins (2001) e Uchôa (2013). Quanto ao nosso entendimento de Estilística, a ciência da linguagem que estuda os estilos das comunicações, nos valem principalmente das teorias de Amado Alonso (*apud* Martins 2012). Assim como Alonso, consideramos indispensável nos valermos inicialmente da Estilística da Língua, de Bally, para entendermos como cada obra é composta de acordo com os aspectos pessoais do autor, demonstrados respeitando as normas gramaticais. Em seguida, deve-se utilizar a Estilística Literária, de Spitzer, para se observar os aspectos psicológicos do autor que interferiram na produção da obra.

Quanto à metodologia utilizada nesta proposta, defendemos plenamente que a LC se mostrou indispensável. Tendo a LC como metodologia e o WST e o índice Flesch como ferramentas de pesquisa, corroboramos Berber Sardinha (1999, 2000, 2008, 2009), quanto a termos acesso a análises confiáveis e comprováveis, em um prazo praticamente impossível de ser obtido caso as análises fossem feitas de forma manual. Entretanto, nos valem deste espaço para registrarmos o quanto a montagem de um *corpus* é laboriosa e requer uma

excessiva dedicação do seu idealizador, conforme demonstramos nesta pesquisa, no capítulo “Metodologia de pesquisa”.

Ao unirmos Léxico, Estilística e LC em nossa proposta, nos valemos da EC para realizarmos as análises literárias e, principalmente, das colocações de Mahlberg (2020). Ao conectar estudos da linguagem com a literatura pela metodologia da LC, a EC tornou possível analisarmos itens como densidade lexical, CT e campos lexicais, entre outros aspectos.

Quanto ao nosso objeto de estudo, o estilo lexical de Paulo Coelho presente em seus romances literários, tendo a LC como metodologia de pesquisa, ressaltamos que a utilização de tal produção literária ocorreu diante da expressividade do autor no campo da literatura. Destacamos que nossas colocações quanto à sua obra são científicas e não retratam nossa opinião pessoal e crítica no sentido literário.

Ao compararmos os dados apresentados pelo WST, entre *O alquimista*, de Paulo Coelho, e 17 obras famosas da literatura brasileira, obtivemos informações relevantes. Destacamos que a escolha por *O alquimista* se deu por ser esta a obra de maior sucesso de Paulo Coelho.

Quanto à STTR, *O alquimista* é a segunda obra em ordem crescente de densidade lexical no conjunto de todas as obras, demonstrando uma baixa densidade entre as palavras utilizadas pelo autor. Entretanto, quando comparamos os romances, encontra-se em padrão semelhante a *Iracema*, de José de Alencar; e *Dom Casmurro*, de Machado de Assis, o que lhe confere uma proximidade a obras clássicas da literatura brasileira no tocante ao aspecto da quantidade de vocabulário.

De acordo com a CT obtida pelos cálculos apresentados pelo Nilc-Matrix, da USP, constatamos que *O alquimista* é de fácil compreensão, porém, ficando em 14º lugar em uma ordem crescente de CT entre as 17 obras. A obra de Paulo Coelho é considerada mais fácil apenas que *Broquéis*, de Cruz e Souza; *O cortiço*, de Aluísio Azevedo; *Vidas secas*, de Graciliano Ramos; e *Os sertões*, de Euclides da Cunha. Entretanto, é mais difícil que todas as demais, incluindo clássicos como *Dom Casmurro*, de Machado de Assis; *Grande sertão: veredas*, de Guimarães Rosa; e *Marília de Dirceu*, de Tomás Antônio Gonzaga.

No conjunto dos 16 romances literários, Paulo Coelho apresenta uma densidade lexical, revelada pela sua STTR, semelhante em todas as obras, demonstrando que não houve aumento ou diminuição com o passar dos anos.

Os índices Flesch demonstram que 15 romances literários de Paulo Coelho apresentam fácil CT e apenas um é considerado difícil (*Hippie*).

Quanto ao uso de *hapax legomena*, mesmo tendo uma densidade lexical considerada baixa quando comparado a outros autores, Paulo Coelho utiliza 32.793 palavras diferentes, sendo que 13.505 são *hapax legomena*, ou seja, 41,18%. Concluimos que, no conjunto das obras de Paulo Coelho, é grande a porcentagem de palavra que aparecem apenas uma vez.

Ao observarmos os campos lexicais apresentados pelas palavra-chave, neste caso substantivos e adjetivos, utilizadas por Paulo Coelho, constamos cinco campos lexicais, com destaque para “Religiosidades e crenças” e “Geografia”. Para isto, o autor se vale de palavras como *Deus*, *anjo*, *milagres* e *paraiso*, além de palavras referentes especificamente a determinadas culturas e religiosidades, como *Aleph* e *Zahir*. No aspecto geográfico, Paulo Coelho utiliza palavras que se referem principalmente a locais na Europa e na Ásia.

Ao contrastarmos as palavras *Deus* e *demônio* no conjunto dos romances de Paulo Coelho, percebemos que o autor remete seus leitores a diferentes concepções de cada ser. *Deus* é bom, justo, paciente e pacificador, além de existir por si mesmo. Ao contrário, *demônio* é ruim, negativo e só se manifesta diante da fragilidade psicológica das pessoas.

Após estas análises, concluimos que o estilo lexical de Paulo Coelho conta com características próprias capazes de explicar o seu sucesso literário. Percebemos que o autor apresenta uma boa acessibilidade textual, tornando suas obras fáceis de serem compreendidas pela maioria dos leitores. Além disso, o autor faz uso de palavras de aspectos religiosos e geográficos que remetem a diversas religiões, culturas e locais existentes principalmente na Europa e na Ásia. Desta forma, o discurso de Paulo Coelho é marcado por um léxico e um estilo que remete os leitores a obras de fácil compreensão e com referências a diversos lugares.

Diante do exposto, entendemos que os objetivos de nossa proposta de pesquisa foram alcançados e visualizamos estudos futuros quanto a análises literárias por meio de uma união entre a Linguística e a Literatura. Pretendemos continuar nossa busca por novas metodologias e ferramentas da LC e CT que permitam observar outros aspectos, assim como diversificarmos nossas análises para com outros autores. Desta forma, pretendemos contribuir com a divulgação das possibilidades de uso da LC e com os estudos linguísticos e literários.

REFERÊNCIAS

- ABBADE, C. M. S. A lexicologia e a teoria dos campos lexicais. In: **Anais do XV Congresso Nacional de Linguística e Filologia**. Cadernos do CNLF, volume XV, número 5, p. 1332-1343. Rio de Janeiro. 2011. Disponível em <http://www.filologia.org.br/xv_cnlftomo_2/105.pdf>. Acesso em 29 set. 2019.
- ALENCAR, J. **Iracema**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2029>. Acesso em 03 jun. 2021.
- ALVES, C. **Espumas flutuantes**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1785>. Acesso em 06 jun. 2021.
- ANCHIETA, J. **Auto representado na Festa de São Lourenço**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1853>. Acesso em 03 jun. 2021.
- ANDRADE, M. **Macunaíma: o herói sem nenhum caráter**. Belo Horizonte: Villa Rica Editoras Reunidas Ltda., 1991.
- ASSIS, M. **Dom Casmurro**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2081>. Acesso em 03 jun. 2021.
- AZEVEDO, G. **O cortiço**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2018>. Acesso em 03 jun. 2021.
- AZEVEDO, A. **Lira dos vinte anos**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1732>. Acesso em 06 jun. 2021.
- BARBOSA, J. B.; MARINE, T. C. Variação e mudança lexical no português brasileiro do século XX: Um estudo descritivo-comparativo. **Revista Signótica**, v. 23, n. 2, p. 307-321, jul./dez. 2011. Disponível em: <<https://www.revistas.ufg.br/sig/article/view/17325>>. Acesso em 23 fev. 2019.
- BARBOSA, M. A. Lexicologia, lexicografia, terminologia, terminografia: objeto, métodos, campos de atuação e de cooperação. In: Estudos Linguísticos XXXIX. **Anais de Seminários do Gel**, Franca, 1991, p. 182-189, 1991. Disponível em: <<https://docplayer.com.br/166584939-Lexicologia-lexicografia-terminologia-terminografia-identidade-cientifica-objeto-metodos-campos-de-atuacao-resumo.html>>. Acesso em 02 set. 2020.
- BEILKE, N. S. V. **Pommersche Korpora: uma proposta metodológica para compilação de corpora dialetais**. 2016. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2016. Disponível em: <<https://repositorio.ufu.br/handle/123456789/18022>>. Acesso em 18 mar. 2019.

BERBER SARDINHA, T. Linguística de *Corpus*: Histórico e Problemática. 2000. In: Scielo Brazil. **Delta**, São Paulo, v. 16, n. 2, 2000. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005>. Acesso em 05 ago. 2017.

BERBER SARDINHA, T. **Pesquisa em Linguística de *Corpus* com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

BERBER SARDINHA, T. **Usando WordSmith Tools na investigação da linguagem**. 1999. In: Direct – Em Direção à Linguagem do Trabalho. Disponível em <<http://www2.lael.pucsp.br/direct/DirectPapers40.pdf>>. Acesso em 09 ago. 2017.

BERBER SARDINHA, T; ALMEIDA, G. M. B. A Linguísticas de *Corpus* no Brasil. In: TAGNIN, S. E. O.; VALE, O. A (org). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008.

BÍBLIA. In: **Bíblia Sagrada Edição Pastoral: Antigo e Novo Testamentos**. Tradução de Ivo Storniolo e Euclides Martins Balancin. São Paulo: Paulus, 1990.

BIDERMAN, M. T. C. A estruturação do léxico e a organização do conhecimento. In: **Letras de Hoje**. Porto Alegre, PUCRS, v. 22, n. 4, p. 81-96, dezembro de 1987. Disponível em <<http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/17049/11065>>. Acesso em 27 set. 2019.

BIDERMAN, M. T. C. As ciências do léxico. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2ª ed. Campo Grande, MS: Ed. UFMS, 2001. p. 13-22.

BIDERMAN, M. T. C. Fundamentos da Lexicologia. In: _____. **Teoria Lingüística**: teoria lexical e computacional. 2ª edição. São Paulo: Martins Fontes, 2001, p. 95-213.

BIDERMAN, M. T. C. **Teoria linguística**: linguística quantitativa e computacional. Rio de Janeiro / São Paulo: LTC, 1978.

BILAC, O. **Alma inquieta**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1996>. Acesso em 06 jun. 2021.

BUARQUE, C. **Budapeste**. 2ª ed. São Paulo: Companhia das Letras, 2003.

CAMINHA, P. V. **A Carta**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2003>. Acesso em 03 jun. 2021.

CARDOSO, E. A.; IGNEZ, A. F. A estilística e o discurso literário contemporâneo. **Matraga**, Rio de Janeiro, v. 20, n. 32, jan./jun. 2013, p. 36-53, 2013. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/matruga/article/view/19838>>. Acesso em 19 maio 2021.

COELHO, P. **A bruxa de Portobello**. São Paulo: Editora Planeta do Brasil, 2006.

- COELHO, P. **A espiã**. 1ª ed. São Paulo: Paralela, 2016.
- COELHO, P. **Adultério**. Rio de Janeiro: Sextante, 2014.
- COELHO, P. **As Valkírias**. Rio de Janeiro: Rocco, 1992.
- COELHO, P. **Brida**. 3ª ed. Rio de Janeiro: Rocco, 1990.
- COELHO, P. **Hippie**. São Paulo: Paralela, 2018.
- COELHO, P. **Na margem do rio Piedra eu sentei e chorei**. Rio de Janeiro: Rocco, 1994.
- COELHO, P. **O Aleph**. Rio de Janeiro: Sextante, 2010.
- COELHO, P. **O alquimista**. 14ª ed. Rio de Janeiro: Rocco, 1988.
- COELHO, P. **O demônio e a Srta. Prym**. Rio de Janeiro: Objetiva, 2000.
- COELHO, P. **O diário de um mago**. 26ª ed. Rio de Janeiro: Rocco, 1987.
- COELHO, P. **O Monte Cinco**. Rio de Janeiro: Objetiva, 1996.
- COELHO, P. **O vencedor está só**. Rio de Janeiro: Agir, 2008.
- COELHO, P. **O Zahir**. Rio de Janeiro: Rocco, 2005.
- COELHO, P. **Onze minutos**. Rio de Janeiro: Rocco, 2003.
- COELHO, P. **Veronika decide morrer**. Rio de Janeiro: Objetiva, 1998.
- CUNHA, E. **Os sertões**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1800>. Acesso em 03 jun. 2021.
- DUARTE, P. M. T. Estilística ou estilísticas?. **Revista Philologus**, Rio de Janeiro, ano 12, número 34, jan./abr., 2006. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/26296>>. Acesso em 29 jul. 2020.
- FINATTO, M. J. B. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. **Organon** – Revista do Instituto de Letras da UFRGS, Porto Alegre, volume 25, número 50, 2011. Disponível em: <<https://seer.ufrgs.br/organon/article/view/28340>>. Acesso em 19 jun. 2021.
- FINATTO, M. J. B. MOTTA, E. Terminologia e Acessibilidade: novas demandas e frentes de pesquisa / Terminology and Accessibility: new demands and research fronts. **Revista GTLex**, Uberlândia, v. 2, n. 2, p. 316-356, jan./jun. 2017. Disponível em: <<https://seer.ufu.br/index.php/GTLex/article/view/44063>>. Acesso em 27 fev. 2022.

FIORIN, J. L. A estilística na tradição de língua portuguesa e os enfoques discursivos atuais. **Caplletra – Revista Internacional de Filologia**, n. 29, 2000. P. 37-52. Disponível em: <<https://www.raco.cat/index.php/Caplletra/article/view/285905>>. Acesso em 23 mai. 2021.

FROMM, G. O uso de corpora na análise linguística. **Revista Factus**, São Paulo v. 1, n. 1, 2003. Disponível em: <http://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/fromm_corpora.pdf>. Acesso em 20 ago. 2017.

FROMM, G.; SANTOS, C. G.; GRAMA, D. F; BEILKE, N. S. V. *Wordsmith Tools e Sketch Engine: um estudo analítico-comparativo para pesquisas científicas com uso de corpora / Wordsmith Tools and Sketch Engine: an analytical-comparative study for scientific research with corpora manipulation*. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 28, n. 3, p. 1101-1248, 2020. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/15766>>. Acesso em 20 set 2021.

GONÇALVES, L. B. Linguística de *corpus* e análise literária: o que revelam as palavras-chave. In: TAGNIN, S. E. O.; VALE, O. A. (Org.). **Avanços da linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p.387-405.

GONZAGA, T. A. **Marília de Dirceu**. <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2012>. Acesso em 03 jun. 2021.

GRAMA, D. F. **Uma análise lexicográfica dos elementos coesivos sequenciais do português para a elaboração de uma proposta de definição**: um estudo com base em *corpus*. 2016. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2016. Disponível em: <<https://repositorio.ufu.br/handle/123456789/18084>>. Acesso em 18 mar. 2019.

GUIMARÃES ROSA, J. **Grande sertão: veredas**. 22ª ed. São Paulo: Companhia das Letras, 2019.

LIMA, V. M. A; SANTOS, C. A. C. M; VOGEL, M. J. M. A teoria do campo semântico no desenvolvimento de vocabulários estruturados para a Web Semântica. In: RIBEIRO, F.; CERVEIRA, M. E. (org.). **Informação e/ou Conhecimento: as duas faces de Jano**. Porto: Faculdade de Letras da Universidade do Porto - CETAC.MEDIA. I Congresso ISKO Espanha e Portugal / XI Congresso ISKO España. 2013. p. 313-329. Disponível em: <<https://repositorio-aberto.up.pt/handle/10216/69630>>. Acesso em 7 ago. 2021.

MAHLBERG, M. **Estilística de Corpus: uma ponte entre os estudos linguísticos e literários**. Tradução de Ariel Novodvorski e Raphael Marco Oliveira Carneiro. Fórum Linguístico, Universidade Federal de Santa Catarina, Florianópolis, SC, v. 17, n. 1, p. 4430-4452, 2020. Disponível em: <<https://periodicos.ufsc.br/index.php/forum/article/view/1984-8412.2020v17n1p4430>>. Acesso em 12 jun. 2020.

MARTINS, N. S. **Introdução à Estilística: A expressividade na Língua Portuguesa**. 4. ed. rev. 2. reimpr. São Paulo: Editora da Universidade de São Paulo, 2012.

MATOS, G. **Seleção de obras poéticas**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1827>. Acesso em 06 jun. 2021.

MINISTÉRIO DA EDUCAÇÃO. **Ensino fundamental de nove anos: passo a passo do processo de implantação**. 2ª edição. Brasília, 2009. Disponível em: <http://portal.mec.gov.br/dmdocuments/passo_a_passo_versao_atual_16_setembro.pdf>. Acesso em 4 jul. 2021.

MOREJÓN, J. G; MARTINS, Manoel Dias. O idealismo linguístico e a estilística literária. **Alfa: Revista de Linguística**, Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), v. 11, 2001, p. 151-164. Disponível em: <<https://periodicos.fclar.unesp.br/alfa/article/view/3302>>. Acesso em 19 maio 2021.

NOVODVORSKI, A.; FINATTO, M. J. B. Linguística de *Corpus* no Brasil: uma aventura mais do que adequada. **Letras & Letras**, v. 30, n.2, jul/dez., p.1-16, 2014. Disponível em: <<http://www.seer.ufu.br/index.php/letraseletras/article/view/28516/15799>> Acesso em 31 ago. 2017.

PARAGUASSU, L. B. **Tradução Especializada Acessível (tea)**: revisão do tema e proposta de disciplina para cursos de graduação em tradução. 2018. Dissertação (Mestrado em Estudos da Linguagem) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018. Disponível em: <<https://lume.ufrgs.br/handle/10183/193093>>. Acesso em 19 jun. 2021.

RAMOS, G. **Vidas Secas**. 131ª ed. Rio de Janeiro: Record, 2016.

ROSA, G. **Grande sertão: veredas**. 1ª ed. São Paulo: Companhia de Bolso, 2021.

SANTOS, C. G. **A criatividade nas denominações de operações policiais**: um estudo preliminar de Campos Semânticos a partir de um corpus de jornais mineiros. 2017. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2017. Disponível em: <<https://repositorio.ufu.br/handle/123456789/20924>>. Acesso em 18 jul. 2021.

SCOTT, M. **WordSmith Tools**. Version 6. Liverpool: Lexical Analysis Software, 2012.

SHEPHERD, T. M. G. Informática e literatura: revelando identidades textuais. In: **Revista Texto Digital**. Periodicos.ufsc.br, Universidade Federal de Santa Catarina, 2004. Disponível em <<https://periodicos.ufsc.br/index.php/textodigital/article/view/1272>>. Acesso em 15 dez. 2019.

SHEPHERD, T. M. G.; BERBER SARDINHA, Tony. A rough guide to doing *Corpus Stylistics*. **Matraga**, Rio de Janeiro, 2013. Disponível em: <https://www.researchgate.net/publication/307850009_A_ROUGH_GUIDE_TO_DOING_CORPUS_STYLISTICS>. Acesso em 15 dez. 2019.

SHEPHERD, T. M. G. Panorama da Linguística de *Corpus*. In: SHEPHERD, T. M. G.; BERBER SARDINHA, T.; PINTO, M. V. (org.). **Caminhos da Linguística de Corpus**. Campinas, SP: Mercado das Letras, 2012.

SILVA, R. S. Riqueza lexical como critério de detecção de autoria. In: **Textos Selecionados**. XXIV Encontro Nacional da Associação Portuguesa de Linguística, Lisboa, Centro de Linguística da Universidade do Porto; *Centre for Forensic Linguistics*, Aston University, 2009. p. 575-587. Disponível em: <<https://apl.pt/wp-content/uploads/2017/09/39-Silva.pdf>>. Acesso em 21 jul. 2021.

SOUSA, C.. **Broquéis**. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2069>. Acesso em 06 jun. 2021.

TAGNIN, S. E. O. E a Linguística de *Corpus* vai desbravando novos horizontes. In: FINATTO, M. J. B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (org.). **Linguística de Corpus: perspectivas**. Porto Alegre: Instituto de Letras – UFRGS, 2018. p. 11-15.

TEIXEIRA, E. D. Linguística de corpus no brasil, ontem e hoje: Uma entrevista com Stella Esther Ortweiller Tagnin. **Belas Infieis**. Brasília, volume 6, número 1, p. 163–175, 2017. Disponível em: <<https://periodicos.unb.br/index.php/belasinfeis/article/view/11424>>. Acesso em 21 abr. 2021.

TEIXEIRA, L. G.; OTTAIANO, A. O. Colocações criativas presentes no *corpus* literário paralelo: Memórias Póstumas de Brás Cubas sob o viés da Linguística de *Corpus*. **Gláuks: Revista de Letras e Artes**, Viçosa, jan./jun, vol. 17, n. 01, p. 221-245. 2017. Disponível em: <<https://www.revistaglauks.ufv.br/Glauks/article/view/12>>. Acesso em 06 mar. 2021.

TOLEDO, L. D. L. **Um estudo sobre o uso de vocabulário rico por aprendizes de inglês na sessão oral do teletandem institucional integrado**, 2017. Dissertação de Mestrado em Estudos Linguísticos – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2017. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/151765/toledo_ldl_me_sjrp.pdf?sequence=3>. Acesso em 27 fev. 2022.

UCHÔA, C. E. F. Estudos estilísticos no Brasil. **Matraga**, Rio de Janeiro, v. 20, n. 32, jan./jun., 2013. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/matraga/article/view/19836>>. Acesso em 18 out. 2019.

VIEGAS, A. M. Conceitos de estilística. **Cadernos de Linguística e Teoria da Literatura**, Universidade Federal de Minas Gerais, n. 7, 1982. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/cltl/article/view/9862>>. Acesso em 23 mai. 2021.

VIEIRA, A. **Sermão de Santo Antônio aos peixes**. Disponível em <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=1744>. Acesso em 03 jun. 2021.

VILELA, M. O léxico do português: perspectivação geral. **Filologia e linguística portuguesa**. São Paulo: USP, 1997, n. 1, p. 31-50. Disponível em: <<http://www.revistas.usp.br/flp/article/view/59644>>. Acesso em 21 mar. 20 mar. 2019.

APÊNDICES

Apêndice A – Estatísticas apresentadas pelo *WordSmith Tools* para as obras de diversos autores

Aluísio Azevedo – <i>O cortiço</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	467.349	81.773	81.735		11.178	13,68	50,69

Álvares de Azevedo – <i>Lira dos vinte anos</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	181.975	31.899	31.893		5.480	17,18	47,86

Castro Alves – <i>Espumas flutuantes</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	108.723	18.592	18.582		4.628	24,91	52,72

Chico Buarque – <i>Budapeste</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	237.328	40.461	40.445		8.079	19,98	51,24

Cruz e Souza – <i>Broquéis</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	35.650	5.369	5.369		2.068	38,52	56,66

Euclides da Cunha – <i>Os sertões</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	1.002.740	155.696	154.670		22.092	14,28	56,58

Graciliano Ramos – <i>Vidas secas</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	157.759	25.030	25.030		5.572	22,26	53,76

Gregório de Matos – <i>Seleção de obras poéticas</i>								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	30.677	5.440	5.417		1.823	33,65	50,16

João Guimarães Rosa – *Grande sertão: veredas*

João Guimarães Rosa - Grande sertão veredas.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	1.099.250	189.171	189.162		19.973	10,56	50,09

José de Alencar – *Iracema*

José de Alencar - Iracema.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	125.967	22.499	22.466		3.877	17,26	45,01

Machado de Assis – *Dom Casmurro*

Machado de Assis - Dom Casmurro.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	374.469	66.858	66.807		8.683	13,00	47,36

Mário de Andrade – *Macunaíma*

Mário de Andrade - Macunaíma.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	257.462	43.217	43.185		8.085	18,72	47,65

Olavo Bilac – *Alma inquieta*

Olavo Bilac - Alma inquieta.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	53.686	9.515	9.514		2.495	26,22	47,81

Padre Antônio Vieira – *Sermão de Santo Antônio aos peixes*

Antônio Vieira - Sermão de Santo Antônio.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	65.929	12.104	12.104		2.717	22,45	42,79

Padre José de Anchieta – *Auto representado na Festa de São Lourenço*

José de Anchieta - Auto representado na Festa de São Lourenço.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	41.454	6.881	6.869		2.200	32,03	50,53

Pero Vaz de Caminha – *A carta*

Pero Vaz de Caminha - A carta.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	43.672	8.042	8.033		1.694	21,09	42,61

Tomás Antônio Gonzaga – *Marília de Dirceu*

Tomás Antônio Gonzaga - Marília de Dirceu.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	119.226	21.229	21.216		4.298	20,26	49,98

Apêndice B – Estatísticas apresentadas pelo *WordSmith Tools* para as obras de Paulo Coelho

Paulo Coelho – <i>A bruxa de Portobello</i>								
A bruxa de Portobello.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	389.853	65.294	65.217		8.272	12,68	49,17

Paulo Coelho – <i>A espiã</i>								
A espiã.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	174.164	29.245	29.146		5.622	19,29	50,35

Paulo Coelho – <i>Adultério</i>								
Adultério.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	346.381	59.679	59.586		8.158	13,69	49,63

Paulo Coelho – <i>As valquírias</i>								
As valquírias.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	228.848	38.084	38.068		5.234	13,75	46,41

Paulo Coelho – <i>Brida</i>								
Brida.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	295.853	49.936	49.922		5.870	11,76	45,27

Paulo Coelho – <i>Hippie</i>								
Hippie.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	350.694	58.997	58.952		8.075	13,70	49,27

Paulo Coelho – <i>Na margem do rio Piedra eu sentei e chorei</i>								
Na margem do rio Piedra eu sentei e chorei.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	207.444	35.377	35.359		4.881	13,80	47,07

Paulo Coelho – <i>O Aleph</i>								
O aleph.lst								
File Edit View Compute Settings Windows Help								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	367.725	62.368	62.255		7.799	12,53	48,23

Paulo Coelho – <i>O alquimista</i>								
---	--	--	--	--	--	--	--	--

O alquimista.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	204.474	35.229	35.228		4.285	12,16	42,79

Paulo Coelho – *O demônio e a senhorita Prym*

O demônio e a senhorita Prym.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	247.669	41.770	41.746		6.074	14,55	48,28

Paulo Coelho – *O diário de um mago*

O diário de um mago.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	343.463	59.438	59.395		7.182	12,09	46,02

Paulo Coelho – *O Monte Cinco*

O Monte Cinco.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	239.408	40.052	40.047		5.811	14,51	47,88

Paulo Coelho – *O vencedor está só*

O vencedor está só.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	649.876	109.270	109.002		12.142	11,14	51,39

Paulo Coelho – *O zahir*

O zahir.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	440.892	74.538	74.428		8.538	11,47	47,94

Paulo Coelho – *Onze minutos*

Onze minutos.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	360.472	61.176	61.153		7.291	11,92	47,56

Paulo Coelho – *Veronika decide morrer*

Veronika decide morrer.lst								
N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	280.558	46.580	46.534		6.639	14,27	48,76

Paulo Coelho – *Todos os 16 romances*

N	text file	file size	tokens (running words) in text	tokens used for word list	sum of entries	types (distinct words)	type/token ratio (TTR)	standardised TTR
1	Overall	5.127.774	867.033	866.038		32.793	3,79	48,21
2	A bruxa de Portobello.bt	389.853	65.294	65.217		8.272	12,68	49,17
3	A espiã.bt	174.164	29.245	29.146		5.622	19,29	50,35
4	Adultério.bt	346.381	59.679	59.586		8.158	13,69	49,63
5	As valquírias.bt	228.848	38.084	38.068		5.234	13,75	46,41
6	Brida.bt	295.853	49.936	49.922		5.870	11,76	45,27
7	Hippie.bt	350.694	58.997	58.952		8.075	13,70	49,27
8	Na margem do rio Piedra eu sentei e chorei.bt	207.444	35.377	35.359		4.881	13,80	47,07
9	O aleph.bt	367.725	62.368	62.255		7.799	12,53	48,23
10	O alquimista.bt	204.474	35.229	35.228		4.285	12,16	42,79
11	O demônio e a senhora Pym.bt	247.669	41.770	41.746		6.074	14,55	48,28
12	O diário de um mago.bt	343.463	59.438	59.395		7.182	12,09	46,02
13	O Monte Cinco.bt	239.408	40.052	40.047		5.811	14,51	47,88
14	O vencedor está só.bt	649.876	109.270	109.002		12.142	11,14	51,39
15	O zahir.bt	440.892	74.538	74.428		8.538	11,47	47,94
16	Onze minutos.bt	360.472	61.176	61.153		7.291	11,92	47,56
17	Veronica decide morrer.bt	280.558	46.580	46.534		6.639	14,27	48,76

Apêndice C – Índices Flesch apresentados para as 17 obras de autores diversos

Alúcio Azevedo – <i>O cortiço</i>					
198	Índices de Leiturabilidade	flesch		Índice Flesch	51.45907
198	Índices de Leiturabilidade	flesch		Índice Flesch	55.71261
198	Índices de Leiturabilidade	flesch		Índice Flesch	60.68952
Índice médio – 55.95373					

Álvares de Azevedo – <i>Lira dos vinte anos</i>					
198	Índices de Leiturabilidade	flesch		Índice Flesch	87.13162
198	Índices de Leiturabilidade	flesch		Índice Flesch	76.84454
198	Índices de Leiturabilidade	flesch		Índice Flesch	76.99547
Índice médio – 80,32387					

Castro Alves – <i>Espumas flutuantes</i>					
198	Índices de Leiturabilidade	flesch		Índice Flesch	81.84625
198	Índices de Leiturabilidade	flesch		Índice Flesch	77.50897
198	Índices de Leiturabilidade	flesch		Índice Flesch	75.2849
Índice médio – 78,21337					

Chico Buarque – <i>Budapeste</i>					
198	Índices de Leiturabilidade	flesch		Índice Flesch	65.62294
198	Índices de Leiturabilidade	flesch		Índice Flesch	67.4721
198	Índices de Leiturabilidade	flesch		Índice Flesch	68.54388
Índice médio – 67,212973					

Cruz e Souza – Broquéis				
198	Índices de Leiturabilidade	flesch	Índice Flesch	57.99135
198	Índices de Leiturabilidade	flesch	Índice Flesch	57.17609
198	Índices de Leiturabilidade	flesch	Índice Flesch	56.33341
				Índice médio – 57,16695

Euclides da Cunha – Os sertões				
198	Índices de Leiturabilidade	flesch	Índice Flesch	18.81746
198	Índices de Leiturabilidade	flesch	Índice Flesch	36.42558
198	Índices de Leiturabilidade	flesch	Índice Flesch	35.69537
				Índice médio – 30,31280

Graciliano Ramos – Vidas secas				
198	Índices de Leiturabilidade	flesch	Índice Flesch	52.88041
198	Índices de Leiturabilidade	flesch	Índice Flesch	52.27775
198	Índices de Leiturabilidade	flesch	Índice Flesch	45.67288
				Índice médio – 50,27701

Gregório de Matos – Seleção de obras poéticas				
198	Índices de Leiturabilidade	flesch	Índice Flesch	79.64069
198	Índices de Leiturabilidade	flesch	Índice Flesch	81.58393
198	Índices de Leiturabilidade	flesch	Índice Flesch	78.94395
				Índice médio – 80,05619

João Guimarães Rosa – Grande sertão: veredas				
198	Índices de Leiturabilidade	flesch	Índice Flesch	81.96064
198	Índices de Leiturabilidade	flesch	Índice Flesch	75.13744
198	Índices de Leiturabilidade	flesch	Índice Flesch	79.17017
				Índice médio – 78,75608

José de Alencar – Iracema				
198	Índices de Leiturabilidade	flesch	Índice Flesch	69.13328
198	Índices de Leiturabilidade	flesch	Índice Flesch	72.28269

198	Índices de Leiturabilidade	flesch	Índice Flesch	67.59438
				Índice médio – 69,67011

Machado de Assis – <i>Dom Casmurro</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	70.94817
198	Índices de Leiturabilidade	flesch	Índice Flesch	63.08863
198	Índices de Leiturabilidade	flesch	Índice Flesch	67.44118
				Índice médio – 67,15932

Mário de Andrade – <i>Macunaíma</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	67.63544
198	Índices de Leiturabilidade	flesch	Índice Flesch	62.74378
198	Índices de Leiturabilidade	flesch	Índice Flesch	71.40104
				Índice médio – 67,26008

Olavo Bilac – <i>Alma inquieta</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	79.42638
198	Índices de Leiturabilidade	flesch	Índice Flesch	81.84767
198	Índices de Leiturabilidade	flesch	Índice Flesch	81.5596
				Índice médio – 80,94455

Padre Antônio Vieira – <i>Sermão de Santo Antônio aos peixes</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	72.74706
198	Índices de Leiturabilidade	flesch	Índice Flesch	65.09304
198	Índices de Leiturabilidade	flesch	Índice Flesch	63.7188
				Índice médio – 67,1863

Padre José de Anchieta – <i>Auto representado na Festa de São Lourenço</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	71.7709
198	Índices de Leiturabilidade	flesch	Índice Flesch	72.42418
198	Índices de Leiturabilidade	flesch	Índice Flesch	85.4944
				Índice médio – 76,56316

Pero Vaz de Caminha – <i>A carta</i>				
---	--	--	--	--

198	Índices de Leiturabilidade	flesch	Índice Flesch	70.84414
198	Índices de Leiturabilidade	flesch	Índice Flesch	70.27088
198	Índices de Leiturabilidade	flesch	Índice Flesch	69.67231
Índice médio – 70,26244				

Tomás Antônio Gonzaga – <i>Marília de Dirceu</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	84.76442
198	Índices de Leiturabilidade	flesch	Índice Flesch	86.43459
198	Índices de Leiturabilidade	flesch	Índice Flesch	81.29641
Índice médio – 84,16514				

Apêndice D – Índices Flesch apresentados para as obras de Paulo Coelho

Paulo Coelho – <i>A bruxa de Portobello</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	50.29572
198	Índices de Leiturabilidade	flesch	Índice Flesch	67.96015
198	Índices de Leiturabilidade	flesch	Índice Flesch	47.78514
Índice médio – 55,34700				

Paulo Coelho – <i>A espiã</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	55.44151
198	Índices de Leiturabilidade	flesch	Índice Flesch	59.00522
198	Índices de Leiturabilidade	flesch	Índice Flesch	48.63425
Índice médio – 54,36032				

Paulo Coelho – <i>Adultério</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	60.82718
198	Índices de Leiturabilidade	flesch	Índice Flesch	64.68389
198	Índices de Leiturabilidade	flesch	Índice Flesch	69.7514
Índice médio – 65,08749				

Paulo Coelho – <i>As valquírias</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	60.93632

198	Índices de Leiturabilidade	flesch	Índice Flesch	60.2047
198	Índices de Leiturabilidade	flesch	Índice Flesch	65.23872
Índice médio – 62,12658				

Paulo Coelho – *Brida*

198	Índices de Leiturabilidade	flesch	Índice Flesch	64.08274
198	Índices de Leiturabilidade	flesch	Índice Flesch	64.9751
198	Índices de Leiturabilidade	flesch	Índice Flesch	72.18716
Índice médio – 67,08166				

Paulo Coelho – *Hippie*

198	Índices de Leiturabilidade	flesch	Índice Flesch	33.65964
198	Índices de Leiturabilidade	flesch	Índice Flesch	54.84875
198	Índices de Leiturabilidade	flesch	Índice Flesch	60.15519
Índice médio – 49,55452				

Paulo Coelho – *Na margem do rio Piedra eu sentei e chorei*

198	Índices de Leiturabilidade	flesch	Índice Flesch	70.85655
198	Índices de Leiturabilidade	flesch	Índice Flesch	70.44295
198	Índices de Leiturabilidade	flesch	Índice Flesch	76.21703
Índice médio – 72,50551				

Paulo Coelho – *O Aleph*

198	Índices de Leiturabilidade	flesch	Índice Flesch	68.27174
198	Índices de Leiturabilidade	flesch	Índice Flesch	63.48507
198	Índices de Leiturabilidade	flesch	Índice Flesch	68.56338
Índice médio – 66,77339				

Paulo Coelho – *O alquimista*

198	Índices de Leiturabilidade	flesch	Índice Flesch	62.52361
198	Índices de Leiturabilidade	flesch	Índice Flesch	56.2718
198	Índices de Leiturabilidade	flesch	Índice Flesch	70.76504
Índice médio – 63,18681				

Paulo Coelho – O demônio e a senhorita Prym				
198	Índices de Leiturabilidade	flesch	Índice Flesch	54.60743
198	Índices de Leiturabilidade	flesch	Índice Flesch	62.22946
198	Índices de Leiturabilidade	flesch	Índice Flesch	65.19308
				Índice médio – 60,67665

Paulo Coelho – O diário de um mago				
198	Índices de Leiturabilidade	flesch	Índice Flesch	54.14699
198	Índices de Leiturabilidade	flesch	Índice Flesch	65.29085
198	Índices de Leiturabilidade	flesch	Índice Flesch	66.4214
				Índice médio – 61,95308

Paulo Coelho – O Monte Cinco				
198	Índices de Leiturabilidade	flesch	Índice Flesch	63.77401
198	Índices de Leiturabilidade	flesch	Índice Flesch	61.95511
198	Índices de Leiturabilidade	flesch	Índice Flesch	67.21731
				Índice médio – 64,31547

Paulo Coelho – O vencedor está só				
198	Índices de Leiturabilidade	flesch	Índice Flesch	55.1123
198	Índices de Leiturabilidade	flesch	Índice Flesch	41.99057
198	Índices de Leiturabilidade	flesch	Índice Flesch	64.85495
				Índice médio – 53,98594

Paulo Coelho – O zahir				
198	Índices de Leiturabilidade	flesch	Índice Flesch	54.27392
198	Índices de Leiturabilidade	flesch	Índice Flesch	70.49561
198	Índices de Leiturabilidade	flesch	Índice Flesch	66.55673
				Índice médio – 63,77542

Paulo Coelho – Onze minutos				
198	Índices de Leiturabilidade	flesch	Índice Flesch	52.44022

198	Índices de Leiturabilidade	flesch	Índice Flesch	58.19722
198	Índices de Leiturabilidade	flesch	Índice Flesch	58.19722
				Índice médio – 56,27822

Paulo Coelho – <i>Veronika decide morrer</i>				
198	Índices de Leiturabilidade	flesch	Índice Flesch	47.4088
198	Índices de Leiturabilidade	flesch	Índice Flesch	47.36406
198	Índices de Leiturabilidade	flesch	Índice Flesch	59.37308
				Índice médio – 51,38198