


O léxico de Paulo Coelho presente em seus romances literários: um estudo baseado em *corpus*

Paulo Coelho's lexicon presented in his literary novels: a corpus-based study

Marden Aleandro Rangel* 

RESUMO: Em nossa pesquisa de mestrado (Rangel, 2022) desenvolvemos uma proposta metodológica capaz de realizar análises linguísticas de obras literárias através da Linguística de *Corpus*. Neste artigo apresentamos um recorte de tal pesquisa com o objetivo principal de analisar o léxico utilizado pelo escritor Paulo Coelho em um *corpus* com 16 dos seus romances literários lançados até a realização da pesquisa e utilizado na mesma. Para isto, nos baseamos em pressupostos teóricos sobre Léxico e Linguística de *Corpus*, sendo esta também a metodologia utilizada. Entre os referenciais teóricos constam, entre outros, Abbade (2011), Biderman (1987 e 2001), Berber Sardinha (1999, 2000 e 2009) e Fromm (2003). Para utilizarmos a Linguística de *Corpus* como metodologia, primeiramente digitalizamos as obras em aparelhos *scanner*, em seguida convertemos do formato PDF para Word, conferimos e corrigimos manualmente os textos e analisamos pelo programa *WordSmith Tools 6.0* (Scott, 2012) e índice Flesch. Os resultados permitiram observar aspectos lexicais de Paulo Coelho, como diversidade lexical, uso específico de determinadas palavras e campos e subcampos lexicais.

ABSTRACT: In our master's research (Rangel, 2022) we developed a methodological proposal capable of carrying out linguistic analyzes of literary works through Corpus Linguistics. In this article we present an excerpt from such research with the main objective of analyzing the lexicon used by the writer Paulo Coelho in a corpus with 16 of his literary novels released until the research was carried out and used in it. For this, we are based on theoretical assumptions about Lexicon and Corpus Linguistics, which is also the methodology used. Theoretical references include, among others, Abbade (2011), Biderman (1987 and 2001), Berber Sardinha (1999, 2000 and 2009) and Fromm (2003). To use Corpus Linguistics as a methodology, we first digitized the works on scanner devices, then converted them from PDF to Word format, manually checked and corrected the texts and analyzed them using the *WordSmith Tools 6.0* program (Scott, 2012) and the Flesch index. The results made it possible to observe Paulo Coelho's lexical aspects, such as lexical diversity, specific use of certain words and lexical fields and subfields.

* Doutorando. Instituto de Letras e Linguística, Programa de Pós-graduação em Estudos Linguísticos (UFU). mardenrangel@gmail.com

PALAVRAS-CHAVE: Léxico. Linguística de *Corpus*. Paulo Coelho.

KEYWORDS: Lexicon. Corpus Linguistics. Paulo Coelho.

1 Introdução

O objetivo deste artigo é analisar o léxico utilizado pelo escritor Paulo Coelho em 16 de seus romances literários quanto a aspectos como diversidade lexical, palavras-chave, uso de *hapax legomena* e campos e subcampos lexicais. O material é um recorte de nossa pesquisa de mestrado (Rangel, 2022), na qual desenvolvemos uma metodologia de análises linguísticas de obras literárias através da Linguística de *Corpus*, tendo tais romances de Paulo Coelho, lançados até a realização da pesquisa, como *corpus* de estudo.

Especificamente neste artigo, utilizamos teorias advindas da pesquisa de mestrado (Rangel, 2022) e referentes às áreas do Léxico e da Linguística de *Corpus*, sendo esta também usada como metodologia.

Quanto a Paulo Coelho, é um dos escritores mais famosos da atualidade, porém não tem boa consideração por parte dos acadêmicos e dos críticos literários. Maicon Tenfen, em sua coluna na revista *Veja*, publicada em agosto de 2017 e intitulada *Aos 70 anos, Paulo Coelho ainda é o nosso maior péssimo escritor*¹, menciona que:

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 (Tenfen, 2017).

¹ Disponível em: <https://veja.abril.com.br/coluna/o-leitor/aos-70-anos-paulo-coelho-ainda-o-nosso-maior-pessimo-escritor>. Acesso em: 03 set. 2023.

No entanto, é quase impossível não mencionarmos Paulo Coelho ao nos referirmos à literatura contemporânea. O escritor iniciou suas produções literárias em 1982, mas obteve sucesso editorial a partir do biênio 1987/1988 com os lançamentos, respectivamente, de *O diário de um mago* e *O alquimista*². Desde então, seus livros foram traduzidos para 88 idiomas e vendidos em mais de 170 países³, totalizando mais de 320 milhões de exemplares comercializados⁴.

O maior sucesso editorial de Paulo Coelho, *O alquimista*, chegou à primeira colocação na lista das obras mais vendidas em 18 países⁵. Até setembro de 2020 a produção literária havia vendido mais de 65 milhões de exemplares em todo o mundo e em diversos idiomas, sendo que cada uma das 45 mil palavras, apenas da versão em inglês, havia rendido ao seu criador a quantia de aproximadamente R\$ 9 mil, totalizando quase R\$ 400 milhões em obras completas⁶. No ano de 2021, *O alquimista* se tornou o livro mais traduzido na história da literatura brasileira e nono na literatura mundial⁷ e até agosto de 2022 constava há 650 semanas na lista de mais vendidos do jornal USA Today⁸.

Mesmo posicionado na linha divisória entre o sucesso e a falta de reconhecimento, Paulo Coelho tem inúmeros leitores de seus livros. Diante disso, é de se imaginar que ele conte com recursos capazes de despertar o interesse das pessoas, sendo o léxico um deles.

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

³ Disponível em: <https://www.metropoles.com/entretenimento/literatura/paulo-coelho-comemora-650-semanas-na-lista-dos-mais-vendidos-do-usa>. Acesso em: 23 jun. 2023.

⁴ Disponível em: <https://www.uol.com.br/splash/noticias/2022/08/24/paulo-coelho-75-anos-principais-obras-do-autor-mais-traduzido-do-mundo.htm>. Acesso em: 23 jun. 2023.

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

⁶ Disponível em: <https://glamurama.uol.com.br/notas/cada-uma-das-45-mil-palavras-de-o-alquimista-rendeu-perto-de-r-9-mil-a-paulo-coelho/>. Acesso em: 23 jun. 2023.

⁷ 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.

⁸ Disponível em: <https://www.metropoles.com/entretenimento/literatura/paulo-coelho-comemora-650-semanas-na-lista-dos-mais-vendidos-do-usa>. Acesso em: 23 jun. 2023.

Quanto ao *corpus* utilizado na pesquisa de mestrado (Rangel, 2022) e neste artigo, o material é composto por 16 romances literários de Paulo Coelho, publicados até a realização da pesquisa e apresentados no quadro 1, a seguir, em ordem de lançamento e com suas respectivas editoras.

Quadro 1 – *Corpus* utilizado e composto pelos 16 romances literários do escritor Paulo Coelho.

Título	Ano de lançamento	Editora
<i>O diário de um mago</i>	1987	Rocco
<i>O alquimista</i>	1988	Rocco
<i>Brida</i>	1990	Rocco
<i>As Valkírias</i>	1992	Rocco
<i>Na margem do rio Piedra eu sentei e chorei</i>	1994	Rocco
<i>O Monte Cinco</i>	1996	Objetiva
<i>Verônica decide morrer</i>	1998	Objetiva
<i>O demônio e a srta. Prym</i>	2000	Objetiva
<i>Onze minutos</i>	2003	Rocco
<i>O Zahir</i>	2005	Rocco
<i>A bruxa de Portobello</i>	2006	Planeta
<i>O vencedor está só</i>	2008	Agir
<i>O Aleph</i>	2010	Sextante
<i>Adultério</i>	2014	Sextante
<i>A espiã</i>	2016	Paralela
<i>Hippie</i>	2018	Paralela

Fonte: elaboração própria.

Este *corpus* é também um recorte do *corpus* utilizado na pesquisa de mestrado (Rangel, 2022), uma vez que também fizemos uso de um segundo *corpus* composto por 17 obras de diferentes autores e movimentos literários.

2 Referenciais teóricos

Nesta seção apresentamos os referenciais teóricos sobre Léxico e Linguística de *Corpus* adotados na elaboração deste artigo.

2.1 Léxico

O léxico de uma língua natural é fundamental ao homem, pois permite a eficácia na transmissão, recepção e perpetuação do conhecimento. Ao conhecer algo, as pessoas assimilam tal novidade através de uma rotulação empregada por uma comunidade linguística e indispensável para que a comunicação ocorra de forma unificada. É diante desta necessidade de comunicação, assim como de transmissão e aquisição do conhecimento, que surge o léxico de uma língua natural, ou seja, o conjunto comum de palavras usadas por um grupo de pessoas para se comunicarem de forma inteligível. Conforme Biderman (2001), o léxico contribui com este processo de conhecimento, permitindo entender como as pessoas registram, codificam, interiorizam e classificam novos saberes a partir de suas realidades vividas.

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).

Em dois textos nos quais aborda o assunto, Biderman (1987 e 2001) considera essa etapa como uma engenhosa estratégia de simbolização, surgida 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 (Biderman, 2001, p. 14).

Entre as principais características do léxico, destaca-se que ele é heterogêneo em seu constante processo de formação. Segundo Vilela (1997), a sua constituição é democrática, não se faz hermeticamente e não é permitida apenas para determinadas categorias, como a dos linguistas e dos escritores. Pelo contrário, ocorre uma união inconsciente entre especialistas e pessoas de diversas classes sociais, profissões, religiões etc. O léxico ainda possui ambiguidades, como arcaísmos convivendo com neologismos e linguagem formal utilizada em conjunto com coloquialismos.

De acordo com Biderman (2001), outro fator importante quanto ao léxico é a sua origem, capaz de permitir contatos com características e patrimônios das comunidades linguísticas, como vocabular, cultural e histórico.

Abbade (2011) corrobora Biderman (2001) quanto aos contatos com o léxico de uma língua natural permitirem acesso às características da sociedade que o utiliza e revelar diversos aspectos, como cultura, tradição, religiosidades, crenças, costumes e escolaridades.

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).

Quanto à existência do léxico, ela se deve principalmente 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 sem vestígios de outras já existentes, caracterizando o *thesaurus* como indispensável para o dinamismo lexical.

Sobre este movimento constante do léxico, Rey-Debove (*apud* Biderman, 1987) afirma que um vocabulário é renovado em torno de 10% no prazo de 25 anos e um idioma pode ter mais de 400 mil palavras. Isto ocorre devido a motivos como

expansão, contração, marginalização, desaparecimento e ressurgimento de palavras. No Brasil, a última atualização do *Vocabulário Ortográfico de Língua Portuguesa* (doravante *Volp*), lançada pela ABL em julho de 2021⁹, evidenciou tal teoria ao revelar o acréscimo de 1.160 palavras, como feminicídio, criptomoeda e telemedicina¹⁰.

Segundo Rey Debove (*apud* Biderman, 1987), uma pessoa memoriza cerca de 20 mil palavras e o uso delas é extremamente complexo, ágil e eficaz. Diante disso, Biderman (1987) afirma que a memória individual é um importante mecanismo na compreensão da comunicação. Ao conviver com várias palavras, as pessoas memorizam algumas delas 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 patrimônio lingüístico herdado, o léxico se estrutura segundo padrões hierárquicos e sistemáticos (Biderman, 1987, p. 83).

Sendo assim, de forma análoga a um labirinto com seus diversos caminhos, o indivíduo que deseja se comunicar se vê diante de várias opções lexicais que integram

⁹ 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.

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

sua memória e, de forma inconsciente e instantânea, utiliza a palavras que melhor lhe convém para se expressar.

2.1.2 Algumas definições

Antes de encerrarmos as abordagens sobre léxico, apresentamos algumas expressões utilizadas neste artigo.

2.1.2.1 Campos lexicais

A teoria dos campos lexicais preconiza que as dependências e semelhanças entre as palavras resultam em seus agrupamentos e utilizações. De acordo com Abbade (2011), o entendimento do significado de cada palavra de um enunciado se faz observando o todo do qual ela faz parte.

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 (Abbate, 2011, p. 1332).

Em Abbade (2011) tomamos conhecimento que, embora exista oposição entre as palavras, os seus significados no enunciado se dependem mutuamente, pois é necessário considerar todo o enunciado para o seu completo entendimento.

2.1.2.2 Palavras gramaticais e lexicais

Inicialmente, utilizamos Abbade (2011, p. 1.333) para definir que palavra “é um termo genérico, tradicionalmente utilizado na língua, fazendo parte do vocabulário de todos os falantes”.

Segundo Martins (2012), as palavras se dividem em gramaticais ou lexicais. Quanto às gramaticais, elas são mais frequentes, pouco numerosas e suas significações dependem dos contextos linguísticos nos quais são utilizadas. “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 exemplos de palavras gramaticais, temos os pronomes, os artigos, os numerais e as conjunções.

Quanto às palavras lexicais, de acordo com Martins (2012), elas são autossuficientes quanto aos seus significados e capazes de despertar representações mentais. “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 são em número indeterminável e sua renovação é perene, causando constantes alterações na língua onde estão inseridas. A lista delas abrange substantivos, adjetivos, advérbios e verbos que exprimem ação ou processo mental.

2.1.2.3 Hapax legomena

Segundo Fromm, Santos, Grama e Beilke (2020), *hapax legomena* são palavras que aparecem uma única vez no *corpus*.

2.1.2.4 Palavras-chave

Conforme Berber Sardinha (2009), as palavras-chave são obtidas por meio de análises estatísticas, nas quais se analisa quais são usadas em maior ou menor frequência, realizando uma comparação entre o *corpus* de estudo e o *corpus* de referência. Dessa análise obtêm-se a chavicidade, ou seja, o resultado do quão importante é a palavra no *corpus* de estudo comparado com o *corpus* de referência.

2.2 Linguística de *Corpus*

A Linguística de *Corpus* (doravante LC), conforme Berber Sardinha (2000), é uma metodologia que se utiliza de evidências extraídas do *corpus*, através de recursos da informática, para pesquisar a língua e sua variedade linguística, apontando resultados extremamente confiáveis.

Beilke (2016) corrobora Berber Sardinha (2000) e descreve a LC como uma abordagem metodológica, fundamentada em dados autênticos e que possibilita produzir saberes variados, além de direcionar para investigações de hipóteses não imaginadas, descobertas e comprovações de fatos linguísticos.

Conforme Gonçalves (2008), a LC permite a utilização de um grande número de textos e facilita pesquisas e críticas literárias, ao fornecer aspectos como palavras-chave e linhas de concordância dos textos estudados.

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

Outros usos da LC são citados por Berber Sardinha (2000), como estudos da linguagem, pesquisas linguísticas com fins comerciais e desenvolvimento de técnicas com Processamento de Linguagem Natural (PLN) em conjunto com a Ciência da Computação. De acordo com o autor, as possibilidades oferecidas pela LC 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, montagens de sistemas de reconhecimento de voz e confirmação de hipóteses previamente elaboradas ou não imaginadas.

3 Metodologia

Nesta seção apresentamos as etapas da montagem do *corpus* para a pesquisa de mestrado (Rangel, 2022) e utilizado também neste arquivo.

A montagem do *corpus* teve início em 2019 e foram utilizados os 16 romances literários de Paulo Coelho, mencionados na seção Introdução, fundamentais pela projeção do nome do autor. Destacamos que a escolha não diminuiu a importância de suas demais produções, como composições musicais e textos jornalísticos.

O *corpus* contém somente o texto principal de cada romance e os prólogos (pois, em algumas obras, fazem parte dos enredos). Foram excluídas notas de rodapé e do autor, dedicatórias e demais citações, que não interferem nos enredos e não se configuram como um processo espontâneo de produção da linguagem. Utilizamos as edições dos anos de lançamento das obras por serem mais fidedignas ao modo de escrever de Paulo Coelho e que não tiveram alterações e correções, inclusive devido ao Acordo Ortográfico da Língua Portuguesa de 1990.

Como tais obras não se encontravam disponíveis em domínio público, utilizamos as edições impressas, mencionando suas editoras atribuições quanto aos direitos autorais. Recortamos as lombadas de cada livro, deixando as folhas soltas e digitalizamos página por página na extensão PDF.

Em seguida, convertemos os arquivos para o Word, formato doc, utilizando recursos de endereços eletrônicos como o ILovePDF¹¹ e o Smallpdf¹², que disponibilizavam a tecnologia *Optical Character Recognition* (doravante OCR), capaz de tornar editáveis caracteres de textos e imagens digitais.

Para cada romance realizamos conferências manuais entre textos digitais e impressos e corrigimos as alterações textuais feitas automaticamente pelos

¹¹ Disponível em: https://www.ilovepdf.com/pt/pdf_para_word. Primeiro acesso em: 20 set. 2019.

¹² Disponível em: <https://smallpdf.com/pt/pdf-para-word>. Primeiro acesso em: 20 set. 2019.

conversores. Ainda nessa fase, conferimos as obras pelo recurso Revisão de Texto, do Word.

Na última etapa da preparação do *corpus* convertemos os textos em Word para o formato txt (Bloco de Notas), na codificação ANSI, com a finalidade de processar no programa *WordSmith Tools*.

A seguir, no quadro 2, apresentamos as atividades realizadas na montagem do *corpus* e o tempo empregado em cada tarefa.

Quadro 2 – Montagem do *corpus* com as devidas atividades realizadas e o tempo/esforço empregado em cada etapa.

Atividades realizadas	Tempo/esforço empregados
Aquisição das obras literárias (16)	Seis meses de buscas pela internet e em sebos literários
Digitalização das obras	10 dias para todas as obras, sendo uma média de duas horas diárias de atividades
Conversões de PDF para Word	8 dias para todas as obras com aproximadamente uma hora diária de atividades
Conferências e correções em cada obra	10 dias em cada obra e cerca de quatro horas diárias de atividades
Conferências e correções de 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 conferências em cada obra com aproximadamente uma hora para cada obra, totalizando 48 conferências
Conversões de Word para TXT	1 dia com cerca de duas horas de atividades
Total de horas trabalhadas	686 horas

Fonte: elaborado pelo autor.

Para as análises lexicais da pesquisa, utilizamos como ferramenta digital o programa *WordSmith Tools* (doravante WST), versão 6.0, de Scott (2012).

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

Inicialmente, utilizamos o recurso *WordList* e obtivemos as listas de palavras das obras analisadas. Pelo item *tokens* tivemos conhecimento do total geral de ocorrências de palavras em cada obra, considerando as repetições. Em *types* verificamos o total de palavras em cada obra sem considerar as repetições. Em *standardised type/token ratio* (doravante STTR) foi possível termos uma relação padronizada com a divisão do total de *types* pelo total de *tokens*, multiplicado por 100, para obtermos os valores em porcentagem e conhecermos a densidade lexical de cada obra. Nesse caso, quanto mais elevada a STTR, maior é a diversidade lexical utilizada. Nessa fase, produzimos duas *wordlists* para cada obra e para o conjunto de todas elas, sendo uma sem *stoplist* e outra com uma *stoplist* fornecida pelo orientador e de autoria da doutora em Estudos Linguísticos, pela Universidade Federal de Uberlândia (doravante UFU), Flávia Santos da Silva. A *wordlist* com *stoplist* contava com 995 palavras gramaticais do português, que não tínhamos interesse de analisar, e não apresentava as palavras lexicais, que queríamos manter, sendo elas substantivos e adjetivos. Com a *wordlist* sem *stoplist* tivemos acesso às palavras mais frequentes e às *hapax legomena*.

Em seguida utilizamos a ferramenta *KeyWords*, do WST, selecionamos as palavras que seriam analisadas e comparamos suas frequências com a lista de referência BP 100K, cedida pelo orientador da pesquisa de mestrado (Rangel, 2022) e escolhida devido ao seu conteúdo com 100 mil primeiras palavras mais frequentes e provenientes 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)¹³. Essa etapa foi realizada para analisarmos as palavras-chave e os campos lexicais dos romances literários de Paulo Coelho.

A terceira etapa de uso do WST consistiu em utilizarmos a ferramenta *Concord* para analisarmos as linhas de concordância dos romances de Paulo Coelho. Criamos duas listas, sendo uma com as ocorrências da palavra *Deus* e outra com as ocorrências da palavra *demônio*, sendo que *Deus* foi mencionado em 920 linhas de concordância e *demônio* em 112. Em seguida, analisamos como o autor utiliza o léxico para definir *Deus* e *demônio* em obras classificadas como sendo de autoajuda. Para isso verificamos os cotextos anteriores e posteriores de cada uso de tais palavras.

4 Análises e resultados

A seguir, apresentamos os resultados obtidos pelas análises dos romances literários de Paulo Coelho.

4.1 Comparativos entre as obras de Paulo Coelho

As análises desta seção foram baseadas em resultados demonstrados pelo WST. A seguir, na tabela 1, apresentamos as comparações entre os romances de Paulo Coelho, listadas em ordem alfabética e com escalas de *tokens*, *types* e STTR.

Tabela 1 – Comparação entre os romances literários de Paulo Coelho em ordem alfabética de título.

Obra	Ano	Tokens	Types	Sttr
<i>A bruxa de Portobello</i>	2006	65.217	8.272	49,17
<i>A espiã</i>	2016	29.146	5.622	50,35
<i>Adultério</i>	2014	59.586	8.158	49,63
<i>As valquírias</i>	1992	38.068	5.234	46,41

¹³ Disponível em <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>. Acesso em: 19 nov. 2021.

<i>Brida</i>	1990	49.922	5.870	45,27
<i>Hippie</i>	2018	58.952	8.075	49,27
<i>Na margem do rio Piedra eu sentei e chorei</i>	1994	35.359	4.881	47,07
<i>O Aleph</i>	2010	62.255	7.799	48,23
<i>O alquimista</i>	1988	35.228	4.285	42,79
<i>O demônio e a senhorita Pryn</i>	2000	41.746	6.074	48,28
<i>O diário de um mago</i>	1987	59.395	7.182	46,02
<i>O monte Cinco</i>	1996	40.047	5.811	47,88
<i>O vencedor está só</i>	2008	109.002	12.142	51,39
<i>O zahir</i>	2005	74.428	8.538	47,94
<i>Onze minutos</i>	2003	61.153	7.291	47,56
<i>Veronika decide morrer</i>	1998	46.534	6.639	48,76

Fonte: elaborada pelo autor.

Pela tabela 1 percebemos que a coluna *tokens* apresenta a obra *A espiã* com o menor número de *tokens* (29.146) e *O vencedor está só* com o maior, sendo 109.002 *tokens*, o que representa uma variação da ordem de 273,98% entre o menor e o maior resultado, explicada devido à maior extensão do texto de *O vencedor está só*.

A obra mais antiga da lista, *O diário de um mago*, de 1987, contém 59.395 *tokens*, enquanto a mais atual, *Hippie*, de 2018, conta com 58.952 *tokens*, revelando um decréscimo de 0,74% na quantidade de *tokens*.

Pela coluna *types* e considerando as datas dos lançamentos das obras, é possível observar que a evolução da variação lexical não segue uma ordem cronológica. 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 lexical da ordem de 99,25%. Incluindo *O vencedor está só*, a variação lexical é de 183,36%.

Pela STTR apresentada, *O alquimista* é a obra com menor diversidade lexical, da ordem de 42,79 e *O vencedor está só* conta com a maior, sendo 51,39. Verificando os anos de lançamento das obras, a mais antiga, *O diário de um mago*, de 1987, tem uma diversidade lexical de 46,02, e a mais recente, *Hippie*, de 2018, apresenta 49,27 de diversidade lexical, um aumento de 7,06%. Entretanto, ao verificarmos a variação dos

resultados, desconsiderando os anos de lançamento, *O alquimista*, de 1988, apresenta a menor diversidade lexical (42,79) e *O vencedor está só*, de 2008, a maior diversidade lexical (51,39), ou seja, uma variação de 20,09%. Com estes dados, percebemos que a diversidade lexical geral utilizada por Paulo Coelho é mediana, apresentando resultado entre 42,79 e 51,39.

4.2 Quantitativo de *types* e *hapax legomena* e análises dos campos lexicais e palavras-chave nos romances literários de Paulo Coelho

Ao utilizarmos as *Wordlists*, produzidas pelo WST, dos romances literários de Paulo Coelho, verificamos que, em um total de 32.793 *types* (palavras diferentes), 13.505 são *hapax legomena*, ou seja, 41,18%. A nossa dedução foi que o escritor conta com uma ampla diversidade lexical ao considerarmos o uso de *hapax legomena*, pois quase metade do seu léxico pertence a esse grupo.

A partir da *Keywords*, do WST, obtivemos as 187 palavras-chave usadas pelo escritor, resultantes de um recorte com substantivos e adjetivos, e as categorizamos em grupos de acordo com definições apresentadas na versão impressa da *Bíblia Sagrada* (1990), no endereço eletrônico Significados¹⁴ e nas versões digitais dos dicionários *Aulete Digital*¹⁵, *Dicio*¹⁶ e *Michaelis*¹⁷. Agrupamos as palavras em campos lexicais relativos às suas áreas de concentração, sendo que algumas eram intercambiáveis e constaram em mais de um campo lexical.

A seguir, apresentamos campos lexicais e tabelas com subcampos lexicais com suas respectivas quantidades de obras nas quais as palavras ocorreram, frequência delas e chavicidade apresentada na coluna *Keyness*, do WST. Todas as tabelas apresentam as palavras em ordem decrescente quanto às suas frequências.

¹⁴ Disponível em: <https://www.significados.com.br>. Acesso em: 18 jul. 2021.

¹⁵ Disponível em: <https://aulete.com.br>. Acesso em: 18 jul. 2021.

¹⁶ Disponível em: <https://www.dicio.com.br>. Acesso em: 18 jul. 2021.

¹⁷ Disponível em: <https://michaelis.uol.com.br>. Acesso em: 18 jul. 2021.

4.2.1 Campo lexical Religiosidades e Crenças

A identificação do campo lexical *Religiosidades e Crenças* resultou na composição de quatro subcampos lexicais e contou com palavras que remetem os leitores a diversas religiões, crenças e divindades. Um dos subcampos é *Divindades*, demonstrado na tabela a seguir.

Tabela 2 – Subcampo: Divindades.

Divindade	Quantidade de obras com ocorrência das palavras	Frequência das palavras	Chavicidade
Deus	16	920	2.499,99
Anjos	16	309	1.894,17
Anjo	16	309	4,176
Deuses	14	140	610,33
Demônio	11	112	2,467,00
Deusa	8	63	325,19

Fonte: elaborada pelo autor.

Nesse subcampo mencionamos a palavra *demônio* pois nas religiões cristãs ele é o anjo que se opôs a Deus. Especificamente, percebemos que as palavras *Deus*, *anjo* e *anjos* são citadas em todos os 16 romances analisados, enquanto *demônio* aparece em 11. *Deus* também é a palavra mais citada em um total de 920 vezes.

A seguir, na tabela 3, apresentamos o subcampo *Locais relacionados às religiosidades*.

Tabela 3 – Subcampo: Locais relacionados às religiosidades.

Locais relacionados às religiosidades	Quantidade de obras com ocorrência das palavras	Frequência das palavras	Chavicidade
Céu	16	267	927,82
Paraíso	15	117	310,07
Céus	16	100	502,11

Fonte: elaborada pelo autor.

A palavra *céu* é a mais mencionada (267 ocorrências) e, juntamente com *céus* (100 ocorrências), as duas são mencionadas em 367 ocorrências.

A palavra *paraíso* (117 ocorrências) consta na tabela 3 por ser mencionada no *Antigo Testamento* como sinônimo de céu e jardim do Éden, criado por Deus para abrigar Adão e Eva; e no *Novo Testamento* como o lugar onde permanecem as almas dos homens bem-aventurados após a morte.

A seguir, na tabela 4, mencionamos o subcampo *Palavras de cunho religioso*.

Tabela 4 – Subcampo: Palavras de cunho religioso.

Palavras de cunho religioso	Quantidade de obras com ocorrência das palavras	Frequência das palavras	Chavicidade
Alma	16	539	2.236,7
Sagrado	14	100	307,66
Perdão	14	91	329,19
Milagres	12	81	338,31
Peregrinos	7	56	338,44

Fonte: elaborada pelo autor.

Na tabela 4 temos a palavra *alma* com 539 ocorrências, mencionada em todas as obras analisadas.

Na tabela 5, a seguir, apresentamos as palavras pertencentes a culturas e locais específicos.

Tabela 5 – Subcampo: Palavras pertencentes a culturas específicas.

Palavras pertencentes a culturas específicas	Quantidade de obras com ocorrência das palavras	Frequência das palavras	Chavicidade
Wicca	1	229	3.063,10
Ágape	1	55	495,14
Aleph	1	51	340,51
Zahir	1	50	451,35

Fonte: elaborada pelo autor.

Conforme verificamos na tabela 5, o subcampo apresenta quatro palavras. Explicamos que *Wicca* diz respeito a uma religião moderna baseada em rituais pagãos antigos, também chamada de bruxaria. A palavra *ágape* é intercambiável, pertencente também ao campo lexical *Sentimentos*, se refere ao amor de Deus para com os homens, sendo também uma refeição compartilhada entre os antigos cristãos. Já *Aleph* é interpretado pelos adeptos das doutrinas cabalísticas como um símbolo místico e espiritual e representa Deus como o começo de tudo. Finalmente, *Zahir* se refere a um conceito ligado à filosofia islâmica, que diz respeito aos comportamentos e atitudes exteriores dos indivíduos.

As palavras *Aleph*, *Wicca*, *Zahir* e *ágape* são citadas em apenas um romance por estarem ligadas ao enredo da obra ou a algum personagem.

4.2.2 Campo lexical *Geografia*

As tabelas a seguir, relativas ao campo lexical *Geografia*, apresentam palavras que se referem a locais específicos ou que fazem parte dos estudos da geografia, em ordem decrescente quanto à frequência das palavras.

Na tabela 6 é apresentado o subcampo *Palavras que remetem a locais específicos*.

Tabela 6 – Subcampo lexical: Palavras que remetem a locais específicos.

Palavras que remetem a locais específicos	Quantidade de obras com ocorrências das palavras	Frequência das palavras	Chavicidade
Santiago	12	182	562,04
Viscos	1	146	1.939,50
Villette	1	75	1.003,10
Assírios	1	72	963,06
Nepal	3	51	300,52
Sidon	1	33	441,40
<i>Lubljana</i>	1	31	414,65
Baikal	2	24	321,02

Fonte: elaborada pelo autor.

Quanto às palavras citadas na tabela 6, explicamos que Santiago é uma região na Espanha (Europa); Viscos é uma cidade na França (Europa), Villette é um sanatório na Eslovênia (Europa), assírios são pessoas pertencentes ou representantes da antiga Assíria (Ásia) e estão na lista por estarem vinculados a tal região, Nepal é um país da Ásia, Sidon é uma cidade no Líbano (Ásia), Lubljana é a capital da Eslovênia (Europa) e Baikal é um lago na Sibéria (Ásia).

Nesse subcampo lexical, os locais são citados, no máximo, em três obras. A única exceção é a região espanhola de Santiago, na Europa, local onde Paulo Coelho peregrinou. Isso agrega uma conotação intimista e autobiográfica do escritor em suas obras. Observamos que todos os locais se encontram na Europa e na Ásia e não são frequentemente mencionados pelos brasileiros.

A tabela 7, a seguir, diz respeito ao subcampo *Locais genéricos*, definidos como de domínio da geografia.

Tabela 7 – Subcampo lexical: Locais genéricos.

Locais genéricos	Quantidade de obras com ocorrências das palavras	Frequência das palavras	Chavicidade
Cidade	16	802	415,06
Deserto	13	373	2.388,30
Montanhas	15	219	1.158,70
Montanha	16	113	458,94
Oásis	3	59	371,06

Fonte: elaborada pelo autor.

Nesse subcampo percebemos que as palavras são citadas em várias obras e certas características cosmopolitas nos romances, ou seja, abarcam uma grande quantidade de lugares.

4.2.3 Campo lexical *Sentimentos*

O campo lexical *Sentimentos* apresenta palavras com definições ligadas a sentimentos.

Tabela 8 – Campo lexical: Sentimentos.

Sentimentos	Quantidade de obras com ocorrências das palavras	Frequência das palavras	Chavicidade
Amor	16	1.340	5.627,90
Medo	16	749	2.598,60
Contente	16	231	1.485,10
Alegria	16	220	621,77
Prazer	15	189	346,62
Feliz	16	187	430,02
Ódio	11	93	308,99
Ágape	1	55	495,14

Fonte: elaboração própria.

Ressaltamos, novamente, que a palavra *ágape* é intercambiável, pertencendo também ao subcampo lexical *Palavras pertencentes a culturas específicas* e significa o amor que Deus tem para com os homens, sendo citada por Paulo Coelho apenas na obra *O diário de um mago*.

Todas as demais palavras deste subcampo são citadas em várias obras, reforçando a perspectiva de obras de autoajuda.

4.2.4 Campo lexical *Anatomia*

As palavras que compõem esse campo lexical são citadas nos dicionários pertencendo à área de anatomia e, geralmente, ao corpo humano.

Tabela 10 – Campo lexical: Anatomia.

Órgãos	Quantidade de obras com ocorrências das palavras	Frequência das palavras	Chavicidade
Olhos	16	908	3.570,20
Coração	16	566	1.850,1
Cabeça	16	559	1,335,30
Mão	16	425	605,99
Mãos	16	412	976,62
Rosto	16	297	1.176,70
Cabelos	16	160	607,94
Pés	16	160	426,46
Braços	16	155	468,65

Fonte: elaborada pelo autor.

Nesse campo lexical, *Anatomia*, todas as palavras são mencionadas em todas as obras.

4.2.5 Campo lexical *Astronomia*

Esse campo lexical, *Astronomia*, apresenta palavras pertencentes à astronomia.

Tabela 11 – Subcampo: Astronomia.

Órgãos	Quantidade de obras com ocorrências das palavras	Frequência das palavras	Chavicidade
Sol	16	448	1.469,20
Universo	16	217	370,75
Lua	13	178	654,47

Fonte: elaborada pelo autor.

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

4.3 Análises das palavras *Deus* e *demônio* e seus cotextos nos romances literários de Paulo Coelho

Ao utilizarmos as listas geradas pela ferramenta *Concord*, do WST, analisamos os cotextos das palavras *Deus* e *demônio*.

4.3.1 *Deus* e seus cotextos nos romances literários de Paulo Coelho

A palavra *Deus* é citada 920 vezes em todos os 16 romances literários de Paulo Coelho, transmitindo uma imagem confiável de *Deus*. A seguir, citamos alguns trechos que aparecem em vários romances, como “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”. A palavra *Deus* também é mencionada no feminino, como em “face feminina de Deus”, “rosto feminino de Deus” e “acredite que Deus é mãe”. Desta forma, percebemos que Paulo Coelho retrata Deus como um ser bom, compassivo e com grande credibilidade pelas pessoas.

Outras características de Deus nos romances de Paulo Coelho o evidencia como o início de tudo, atuante, onipotente, onisciente e onipresente. Apresentamos a seguir alguns trechos que confirmam estas deduções, cujos títulos das obras estão entre parênteses: “Deus aceitou teu combate” (*O Monte Cinco*, 1996), “Deus enviará seu exército de anjos” (*As Valkírias*, 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*) e “Deus ficou contente” (*O demônio e a Srta. Prym*, 2000).

4.3.2 *Demônio e seus cotextos no conjunto nos romances literários de Paulo Coelho*

A palavra *demônio* é citada 112 vezes em 11 dos 16 romances literários de Paulo Coelho. Percebemos que demônio é mostrado como um ser que precisa de alguém fraco psicologicamente para se manifestar, pois ele é consequência de atos e pensamentos nocivos do homem consigo mesmo e ao próximo. Os trechos a seguir confirmam tal constatação, foram retirados dos romances e novamente os títulos das obras estão entre parênteses: “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. Prym*, 2000).

A seguir apresentamos os três trechos mais extensos sobre 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).

4.3.3 Deus e demônio no léxico de Paulo Coelho

Após as análises, concluímos que Paulo Coelho utiliza seu léxico para reforçar conceitos sobre *Deus* e *demônio* que são comuns nas religiões cristãs, não apresentando algo novo.

Assim como nas religiões cristãs, Deus é bom nos romances de Paulo Coelho e existe independente de alguém. Pelo contrário, as pessoas é que precisam de Deus para serem felizes, pois ele é a própria felicidade. Deus também é retratado como um ser justo, milagroso, bondoso, amoroso e sempre disposto a realizar o bem.

Ao contrário, demônio é mostrado como um ser ruim nos romances de Paulo Coelho e precisa das pessoas para se manifestar e realizar suas proezas. Muitas vezes, ele é fruto de pensamentos de pessoas fracas psicologicamente. Caso contrário, diante de alguém firme em sua fé e boas convicções, o demônio perde sua força e não encontra espaço para praticar suas ações.

5 Considerações finais

Ao concluirmos este artigo, destacamos novamente que a escolha pelo *corpus* composto por 16 romances literários de Paulo Coelho se fez pela expressividade do autor no campo da literatura e que nossas colocações são científicas e linguísticas, não retratando opiniões pessoais e críticas literárias.

Em seus romances literários, Paulo Coelho apresenta uma diversidade lexical, revelada pela STTR, semelhante em todas as obras. Isso evidencia que o autor não registrou evolução cronológica em sua diversidade lexical.

Ainda sobre diversidade lexical, Paulo Coelho utiliza 32.793 *types*, sendo que 13.505 são *hapax legomena*, ou seja, 41,18%. No total geral das obras, é grande a porcentagem de palavras usadas apenas uma vez e isso comprova que o léxico utilizado por Paulo Coelho não é repetitivo, evidenciando uma boa diversidade lexical.

Ao observarmos os campos e subcampos lexicais apresentados pelas palavras-chave, neste caso substantivos e adjetivos, utilizadas por Paulo Coelho, constatamos cinco campos mais evidentes. O maior é o campo *Religiosidades e Crenças*, que mostra o quanto o autor se vale de palavras como *Deus*, *anjo*, *milagres* e *paraíso*, além de palavras referentes especificamente a determinadas culturas e religiosidades, como *Aleph* e *Zahir*. Desta forma, as obras remetem seus a reflexões de cunho religioso, autoajuda e autoconhecimento.

Outro campo que merece destaque é o *Geografia*. No aspecto geográfico, Paulo Coelho utiliza palavras que se referem principalmente a locais na Europa e na Ásia, pouco comentados e de conhecimento dos brasileiros.

Ao contrastarmos as palavras *Deus* e *demônio*, percebemos que o autor remete seus leitores a diferentes concepções de cada ser. Nas obras, 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.

Concluimos que o léxico utilizado por Paulo Coelho em seus romances literários é variado, ou seja, conta com uma boa diversidade lexical. O léxico do autor também conta com características próprias, como o uso de palavras de aspectos religiosos e geográficos que remetem a diversas religiões, culturas e locais, existentes principalmente na Europa e na Ásia. Dessa forma, entrega aos leitores obras com referências a diversos lugares.

Por fim, diante do exposto, entendemos que analisamos o léxico de Paulo Coelho e esperamos ter contribuído para futuros estudos sobre léxico, linguística, literatura e produções literárias do referido autor, tendo como metodologia a Linguística de *Corpus*.

Referências

ABBADE, C. M. S. A lexicologia e a teoria dos campos lexicais. **Anais do XV Congresso Nacional de Linguística e Filologia**. Cadernos do CNLF, volume XV, número 5. Rio de Janeiro: 2011. p. 1332-1343. Disponível em: http://www.filologia.org.br/xv_cnlf/tomo_2/105.pdf. Acesso em: 29 set. 2019.

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. 286 f. Disponível em: <https://repositorio.ufu.br/handle/123456789/18022>. Acesso em: 18 mar. 2019.

BERBER SARDINHA, T. Linguística de Corpus: Histórico e Problemática. 2000. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**. São Paulo: v. 16, n. 2, 2000. Acesso em: 05 ago. 2017. DOI <https://doi.org/10.1590/S0102-44502000000200005>

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. **Direct** – Em Direção à Linguagem do Trabalho. Disponível em <http://www2.lael.pucsp.br/direct/DirectPapers40.pdf>. Acesso em: 09 ago. 2017.

BÍBLIA. **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. **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: Ed. UFMS, 2001. p. 13-22.

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

- 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.
- 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. Acesso em: 20 set 2021. DOI <https://doi.org/10.17851/2237-2083.28.3.1191-1248>

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.

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.

RANGEL, M. A. **Proposta metodológica para análises de obras literárias através da Linguística de Corpus: o caso de Paulo Coelho**. 2022. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2022. 91 f. Disponível em: <https://repositorio.ufu.br/handle/123456789/34246>. Acesso em: 07 mar. 2024.

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

TENFEN, M. Aos 70 anos, Paulo Coelho ainda é o nosso maior péssimo escritor. **Veja** [on-line]. São Paulo, 24 ago. 2017. Cultura. Disponível em: <https://veja.abril.com.br/coluna/o-leitor/aos-70-anos-paulo-coelho-ainda-o-nosso-maior-pessimo-escritor>. Acesso em: 03 set. 2023.

VILELA, M. O léxico do português: perspectiva geral. **Filologia e linguística portuguesa**. São Paulo: USP, 1997, n. 1, p. 31-50. Acesso em: 20 mar. 2019. DOI <https://doi.org/10.11606/issn.2176-9419.v0i1p31-50>

Artigo recebido em: 12.12.2023

Artigo aprovado em: 21.03.2024