

# Análise Lexicométrica de *Luanda, Lisboa, Paraíso*, de Djaimilia Pereira de Almeida

<https://doi.org/10.21814/uminho.ed.50.11>

Carla Sofia Araújo

## Introdução

Este artigo sustenta-se na tradição empírica e investigativa da linguística de *corpus* (Biber et al., 1998; Sardinha, 2004; Sinclair, 1991). O trabalho é constituído por duas partes. A primeira parte inclui o enquadramento teórico e contempla a definição de conceitos abordados no trabalho, tais como “lexicometria”, “campo temático” e “palavra-tema”. Na segunda parte, utilizando o Nooj (Silberztein, 2003), efetuaremos uma análise lexicométrica, baseada na análise estatística das palavras-tema, tendo em vista a delimitação de possíveis campos temáticos no romance *Luanda, Lisboa, Paraíso* de Djaimilia Pereira de Almeida (2018). Nesse sentido, começaremos por apresentar os dados gerais do *corpus* e organizar uma listagem de palavras-tema, partindo da listagem dos *tokens* por ordem decrescente de frequência, fornecida automaticamente pelo Nooj.

---

Carla Sofia Araújo, Departamento de Português, Escola Superior de Educação, Instituto Politécnico de Bragança, Bragança, Portugal/Centro de Estudos em Letras, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal <https://orcid.org/0000-0002-5318-0960>  
carla.araujo@ipb.pt

A análise lexicométrica do romance *Luanda, Lisboa, Paraíso*, de Djaimilia Pereira de Almeida, termina com a apresentação dos campos temáticos, delimitados a partir das respetivas palavras-tema.

## Enquadramento Teórico

### Lexicometria

A lexicometria é um processo metodológico e tecnológico de pendor objetivo, descritivo, indutivo e científico. Os procedimentos lexicométricos permitem tratar estatisticamente dados qualitativos sob fundo quantitativo, visando caracterizar o contexto e a combinação dos elementos lexicais de um determinado *corpus* de análise. A lexicometria permite efetuar “reorganizações formais do vocabulário (conjunto de formas atualizadas no discurso, atestadas num texto ou num *corpus* de textos). O estudo lexicométrico impõe o levantamento exaustivo de TODAS as ocorrências, de TODAS as formas do *corpus* a estudar” (Carvalho et al., 1999, p. 225).

De acordo com Leblanc (2015), podemos observar dois tipos de ferramentas lexicométricas: ferramentas lexicométricas tradicionais e ferramentas lexicométricas estruturantes. As primeiras possibilitam a análise dos dados linguísticos caracterizadores dos textos, tais como a frequência das palavras, as formas *hápax legómenon* (formas que ocorrem apenas uma vez no *corpus*), as formas gramaticais, a recorrência de diversos tipos de unidades multilexicais, do ponto de vista da sua estrutura gramatical, entre outros. As segundas permitem análises baseadas em probabilidades.

O levantamento de dados linguísticos quantificados, a sua organização e a análise objetiva revelam fenómenos linguísticos não observáveis nas análises de pendor mais tradicional.

Alguns autores caracterizam a metodologia lexicométrica como uma possível ferramenta de análise de discurso (Leblanc, 2015; Marchand, 2013; Salem, 1986). Os universos de discurso constituem

vertentes de reconhecimento ideológico e, especialmente, de produção de sentido. Com efeito, as ocorrências linguístico-textuais congregam relações de sentido ligadas às manifestações discursivas que podem ser explicitadas através dos dados quantificados. A diversidade discursiva do universo do discurso literário transformá-lo num profícuo ambiente para análises linguísticas, sobretudo no nível lexical.

O texto literário configura um repositório linguístico fértil de produção de sentido, sendo o processo de produção de sentido veiculado através do léxico. Nos variados universos discursivos, podemos constatar uma panóplia diversificada de dimensões socioculturais de uma determinada sociedade. Cada universo de discurso é caracterizado por um determinado padrão que se manifesta nos mais diversos níveis linguísticos, dentre eles o lexical.

O léxico posiciona-se na interseção de outras áreas da linguística, da fonologia e da morfologia no âmbito da forma das palavras, da semântica sobre a sua significação e da sintaxe para as suas propriedades combinatórias. O léxico, ao invés de constituir um sistema, no sentido preciso, é um conjunto aberto e não autónomo. Consequentemente, o léxico não é passível de ser descrito de modo sistemático ou simples, sendo apenas passível de descrições complementares, consoante a perspetiva assumida. Segundo Melčuk et al. (1995), devido à sua vocação transdisciplinar, a lexicologia não apresenta um módulo lexicológico específico, contrariamente à fonética, à morfologia ou à sintaxe, por exemplo. Esta especificidade concede-lhe um papel de interface em relação às diversas áreas disciplinares e níveis de análise em linguística. Deste modo, a lexicologia configura uma ciência “Carrefour” (Niklas-Salminen, 1997, p. 5), na qualidade de disciplina que se centra no estudo das relações sistemáticas que se estabelecem entre as unidades lexicais.

Entre o final dos anos 50 e a década de 70 do século XX, os estudos em lexicologia revelam uma valorização da vertente socio-semântica e sociolinguística do léxico (Tamba-Mecz, 1998). Nas últimas décadas do século XX, assiste-se a uma nova dinâmica em lexicologia, mercê da emergência de novos paradigmas em semântica, da

consolidação da pragmática, e da neutralização da oposição lexical *versus* gramatical.

Assim, tal como defende Rey-Debove (1998), o objeto de estudo da lexicologia situa-se entre o sistema, o uso e o discurso. Niklas-Salminen (1997) preconiza que o léxico é uma entidade teórica e uma realidade da língua, distinta do vocabulário, que se situa no plano discursivo:

o léxico de uma língua deve ser considerado, acima de tudo, como uma entidade teórica. É o conjunto de palavras que uma língua disponibiliza para os falantes ( ... ) o léxico é uma realidade linguística que não pode ser acedida somente pelo conhecimento dos vocabulários particulares que são uma realidade do discurso. (p. 27)

Nesse sentido, efetuar análises lexicais não é tarefa fácil, uma vez que o léxico corresponde ao nível linguístico mais extralinguístico da língua, dado que está ligado às influências culturais, sociais, históricas e ideológicas. Por conseguinte, Vilela (1994) refere que “o léxico é o repositório do saber linguístico e é ainda a janela através da qual um povo vê o mundo” (p. 6). De facto, o léxico é a janela através da qual vemos o mundo e é também a porta de entrada para a leitura de textos. No texto, repousa toda a cultura, da qual o léxico é seu repositório, sendo o léxico o “tesouro vocabular de uma determinada língua” (Biderman, 1981, p. 138).

### Linguística Sistémico-Funcional e Linguística de *Corpus*

Preconizando que a língua é uma das principais revelações da cultura de um povo, a gramática sistémico-funcional de Halliday (1985) focaliza-se no texto, porque é no texto que o discurso se instaura. Ao interpretarmos a língua, podemos verificar que os usos mais correntes nos remetem para a estrutura social, os valores, os sistemas de conhecimento para todos os padrões da cultura. A descrição de uma língua permite-nos aceder a esses sistemas de conhecimento e padrões culturais em que o seu sistema conceptual

se ancora. Considerando o uso como a dimensão essencial para caracterizar e descrever a língua, Halliday (1973) perspectiva a língua como um potencial de significados, um potencial compartilhado, que “é neutro em relação ao falante e ao ouvinte, mas pressupõe falante, ouvinte e situação” (p. 129). O texto representa a atualização do potencial de significado da língua, o processo de escolha semântica.

A multiplicidade de escolhas de significados que a língua nos oferece procede das funções básicas da linguagem. Atendendo aos diversos usos sociais da língua, Halliday (1985) apresenta três metafunções da linguagem: a ideacional, a interpessoal e a textual.

A metafunção ideacional decorre da função representacional da linguagem, está relacionada com a expressão do conteúdo, ligada às potencialidades de significado das expressões linguísticas e às experiências dos mundos interno e externo do falante/ouvinte. Esta metafunção, para além de especificar as escolhas disponíveis no significado, também baliza a natureza das suas realizações estruturais.

A metafunção interpessoal envolve todos os usos da língua para revelar relações sociais e pessoais, bem como todas as formas de intervenção do falante na situação de fala e no ato de fala. Essa metafunção explana o efeito do ato de fala sobre os outros. A interação verbal revela-se com maior evidência nesta metafunção, que nos remete para a função interpessoal da linguagem, todavia a mesma não é predominante relativamente às outras duas metafunções.

Por fim, através da metafunção textual, Halliday (1985) põe em evidência que a linguagem tem uma função textual, pois usamo-la para organizar os significados ideacionais e interpessoais de forma cabal, linear e coerente. Por conseguinte, a metafunção textual concede materialidade às duas funções anteriormente referidas, colmata a exigência de que uma comunicação verbal será coerente e significativa, em contextos situacionais concretos. A ligação de uma frase a outras, a construção dos potenciais de significado e a sua interligação na construção do texto decorrem da metafunção textual. Diz, portanto, respeito à capacidade que o falante possui para construir textos situacionalmente adequados e estruturalmente coesos e coerentes.

Desta forma, o modelo da linguística sistémico-funcional, de base semântica, apresenta duas perspetivas complementares da língua: a língua enquanto sistema e a língua enquanto conjunto de textos, ou seja, de duas perspetivas, a da potencialidade e a da instanciação, respetivamente. Nesse sentido, a linguística sistémico-funcional olha para o texto como um instrumento para o conhecimento do sistema, isto é, o texto constitui uma unidade de análise e de descrição.

Na sequência de pesquisas linguísticas realizadas recentemente no âmbito da linguística sistémico-funcional e da linguística de *corpus*, operaram-se avanços significativos relativamente à conceção da microestrutura e macroestrutura textuais, decorrentes da crescente intenção de qualificar dados quantitativos, nomeadamente as escolhas lexicogramaticais e os padrões de ocorrências nos géneros discursivos. Por conseguinte, a associação das ferramentas metodológicas fornecidas pelo suporte teórico de ambas as teorias permite um estudo abrangente dos textos.

Diversos autores associam a linguística de *corpus* à teoria funcionalista da linguagem desenvolvida por Halliday (1985), a linguística sistémico-funcional. Tal associação decorre do facto da mesma perspetivar a linguagem como um sistema probabilístico de escolhas, através do qual traços linguísticos são determinados.

De acordo com Berber Sardinha (2004),

a história da Linguística de *Corpus* está condicionada à tecnologia, que permite não somente o armazenamento de *corpora*, mas também a sua exploração. Por isso, a história da área está relacionada à disponibilidade de ferramentas computacionais para análise de *corpus*. (p. 5)

Deste modo, a linguística de *corpus* disponibiliza valiosas metodologias para a análise linguística, tal como a que apresentamos na Secção 3 deste trabalho, onde está patente a análise lexicométrica do romance *Luanda, Lisboa, Paraíso*, de Djaimilia Pereira de Almeida.

## Campo Temático e Palavra-Tema

Dado que o objetivo da nossa análise lexicométrica do romance *Luanda, Lisboa, Paraíso* é identificar campos temáticos, a partir da delimitação das palavras-tema na obra de Djaimilia Pereira de Almeida (2018), seguidamente, explanaremos sobre campo temático e palavra-tema.

Galisson e Coste (1976/1983), no *Dicionário de Termos Linguísticos*, apresentam-nos a seguinte definição de campo temático:

os campos temáticos constituem conjuntos de termos funcionalmente possíveis no interior de uma determinada situação temática e cuja organização interna depende de um certo número de parâmetros emprestados à atividade psicossocial. Ex: o campo temático da “casa” compreenderia o que diz respeito ao “edifício” (hall, escada, elevador, degrau, etc.), à “construção” (materiais, etc.), ao “lugar de habitação” (função, decoração, etc.), ( ... ) e a organização destes termos dependeria das atividades do indivíduo que se encontrasse nessa situação temática. (p. 104)

A amplitude do campo temático possibilita que o mesmo campo temático integre, simultaneamente, vários campos lexicais.

Shaw (1976/1982) considera que, na esfera do estudo do texto literário, os campos temáticos configuram a melhor forma de identificar o tema da obra de arte, a ideia principal transmitida.

O conceito de “campo temático” relaciona-se com os conceitos de “palavra-chave” e de “palavra-tema”. No *Dicionário de Termos Linguísticos*, encontramos as seguintes definições:

a palavra-chave é uma palavra plena (não gramatical), de grande frequência numa obra (ou em toda a obra) de um autor; esta frequência apresenta a característica – em relação

à palavra-tema – de estar muito longe da frequência da mesma palavra num corpus de obras do mesmo género. Por outras palavras, a palavra-chave possui a particularidade de ser anormalmente frequente numa obra ou num autor. É classificada de “chave” precisamente por causa desta “anomalia”: ao marcar o desvio entre o que acontece no discurso dos outros e no seu, o autor privilegia a palavra-chave (...). Contrariamente à palavra-tema, o assunto tratado não justifica a elevada frequência da palavra-chave, a sua forte reiteração não se explica senão pela livre escolha do autor: dois escritores que tratem do mesmo tema geralmente fazem aparecer as mesmas palavras-tema, mas apresentam palavras-chave diferentes. (Galissou & Coste, 1976/1983, pp. 114–115)

Genouvrier e Peytard (1974) consideram que a

palavra-tema é uma palavra caracterizada por uma frequência muito elevada e que, numa ordenação por frequência decrescente do vocabulário de um autor, pertence, por exemplo, aos primeiros 50 lugares; palavra-chave é uma palavra cuja frequência apresenta uma diferença máxima (num texto dado) em relação à frequência normal (em outros enunciados). (p. 313)

Para estes autores, “as análises por palavras-tema e palavras-chave permitem caracterizar o estilo do autor como desvio a partir de uma norma (...). Tanto as palavras-chave como hapaxes legomena, isto é os termos exclusivos, são analisados no sentido de caracterizar áreas temático semânticas típicas” (Genouvrier & Peytard, 1974, pp. 317–318).

### **Análise Lexicométrica do Romance *Luanda, Lisboa, Paraíso*, de Djaimília Pereira de Almeida**

Iniciada a análise linguística, o Nooj apresenta-nos os dados gerais caracterizadores do texto, patentes na Tabela 1.

**Tabela 1** Dados gerais do *corpus* de Luanda, Lisboa, Paraíso.

Dados	Frequência
<i>Tokens</i>	52.512
<i>Word forms</i>	44.583
<i>Delimiters</i>	7.583
Anotações	146.629
Ambiguidade	5.494 tipos diferentes de ambiguidade
Unidades linguísticas não ambíguas	2.833

O programa Nooj analisou os *tokens* (cada forma que ocorre no *corpus*) e as respetivas frequências. Os *tokens* podem ser apresentados por ordem decrescente da sua frequência e/ou alfabeticamente.

Através da análise dos itens mais frequentes, verificámos que, tal como ocorre na maioria dos *corpora*, as formas mais frequentes são palavras funcionais ou gramaticais, por exemplo, em *Luanda, Lisboa, Paraíso*, os cinco *tokens* mais frequentes são os seguintes: “De”/“de”, “A”/“a”, “O”/“o”, “Que”/“que” e “E”/“e”, apresentando uma frequência de 1.954, 1.742, 1.439, 1.168, e 1.114, respetivamente.

Na análise dos *tokens* mais frequentes, é importante salientar que o Nooj, tal como se verifica na maioria das aplicações de cariz lexicométrico, procede à distinção entre maiúsculas e minúsculas, considerando, separadamente, cada forma diferente do mesmo lema. Nesse sentido, o *token* que surge no primeiro lugar da lista por ordem decrescente de frequência, “de”, possui uma frequência de 1.932, à qual se soma a frequência da forma “De”, correspondente a 22. De igual modo, o *token* “a” apresenta uma frequência de 1.601, apresentando a forma “A” 141 de frequência.

Selecionados os 60 *tokens* mais frequentes, filtrámos, exportámos os dados e copiámo-los para o Microsoft Word, elaborando uma listagem de palavras-tema e das respetivas frequências, abaixo transcrita (Tabela 2). Optámos por apresentar a listagem de *tokens*, tendo como critérios os nomes comuns, os adjetivos e os verbos principais (Mateus et al., 2003).

**Tabela 2** Palavras-tema de *Luanda, Lisboa, Paraíso*.

Frequência	Token	Frequência	Token	Frequência	Token	Frequência	Token
193	Pai/pai	52	Cama/ cama	38	anos	31	costas
127	Casa/ casa	49	ver	37	Fazia/ fazia	31	criança
123	filho	49	Mãe/ mãe	37	homens	30	cabelo
83	olhos	48	cabeça	37	Vai/vai	29	dizer
82	vez	48	Mulher/ mulher	36	fim	29	meio
82	vida	47	menino	36	porta	29	miúdo
71	dia	46	Fazer/ fazer	35	Rua/rua	29	rapaz
69	homem	45	lado	35	volta	29	saber
68	tempo	42	água	34	boca	28	dedos
59	quarto	42	Sentia/ sentia	34	Velho/ velho	28	Dizia/ dizia
58	Mão/ mão	41	corpo	33	la/ia	28	mesa
58	Papá	41	dar	32	cidade	28	pé
55	noite	41	Dava/ dava	32	vezes	27	cara
53	lugar	40	Sabia/ sabia	31	amigo	27	chão
53	mãos	39	novo	31	calca- nhar	27	marido

Na primeira etapa de análise das palavras-tema, deparamo-nos com fenômenos de ambiguidade potencial, inerente à maior parte das formas linguísticas, uma vez que, na listagem dos *tokens* fornecida pelo Nooj, as palavras-tema não surgem integradas nos próprios contextos de ocorrência, por isso, foi necessário proceder à extração de concordâncias no Nooj. Assim, foram analisados todos os contextos de ocorrência de cada uma das palavras-tema do *corpus* e consultadas as definições patentes no dicionário relativas às mesmas.

The screenshot shows a concordance tool window titled 'CONCORDANCE'. It has a menu bar (File, Edit, Lab, Project, Windows, Info) and a toolbar with options like 'Reset', 'Display', 'characters', 'before, and', '5 after', 'Display', 'Matches', and 'Outputs'. Below the toolbar is a table with three columns: 'Text', 'Before', and 'Seq. After'. The 'Text' column contains a list of sentences. The 'Before' column shows the word 'casa' in red, and the 'Seq. After' column shows the text following the word, also in red. The sentences are:

Text	Before	Seq. After
Aquelas inspirava-lhe dentro de	casa	a solidiedade do trato público
desse começo de vida, muma	casa	térrea onde nunca falara coisa
com as amigas saindo de	casa	aperaltada: a túnica negra lula
Houve que excoitar lá de	casa	quem o tomasse por mau
a criança preferia andar pela	casa	a cavalo na vassoura. Rega
que não tinham gerador em	casa	. O pai era o bocado
a criança. Mantinha-o em	casa	a pretexto de deveres e
o pai passou-os em	casa	. Em oitenta, para seu descontentamento
um soldado ferido volta a	casa	. «Papá, vê só, a gaze
ficar entregue às raparigas da	casa	, cujo crescimento sentia torná-las
sala nem se levavam para	casa	os namorados. Cancelou-se o
sido. Ressentiam o silêncio da	casa	. A luz a que imaginavam
mas estavam surdos. Se em	casa	havia rebuçados para a tosse
ao pescção dele, já a	casa	dormia, deitados na cama tinha
as portadas das janelas da	casa	onde nascera Justina, na zona
mãionese de lagosta em sua	casa	, na zona fina da cidade
a dançarem muma moldura, muma	casa	a óleo, fora do tempo
do sério a dona da	casa	, a quem as ofrendas, engrinaldando
da Cunha à porta da	casa	de pasto. O médico apareceu
tarde, mas ela queria a	casa	a cheirar a canela, o
farófas pela primeira vez em	casa	dos Barbosa da Cunha. Chegaram
cobardia, e metia-se na	casa	de bicho, nauseado, para cuspir
de Glória bailando descalça pela	casa	naquele tempo em Moçamedes, e
rodava a saia estando a	casa	em sossego, aplicava-lhe brilhantina
dividas, antes de sair de	casa	para fazer a chamada, cortava
Lembra quando você chegava a	casa	? Às vezes tocam na porta
à Dona Ehbira do 109, uma	casa	de penhores de touca que
tá aí. Eu mesmo nessa	casa	sem homem, às vezes me

Figura 1 Concordâncias da palavra-tema “casa”.

De facto, a consulta das concordâncias e do dicionário permite um distanciamento relativamente à intuição inerente ao ponto de vista do observador. A título de exemplo, veja-se a palavra-tema “casa”, que configura um fenómeno de homonímia parcial, ou seja, “casa” (verbo casar) e “casa” (nome), conforme se verifica na consulta do dicionário, remetendo-nos para a forma do verbo “casar” e para o nome feminino singular “casa”.

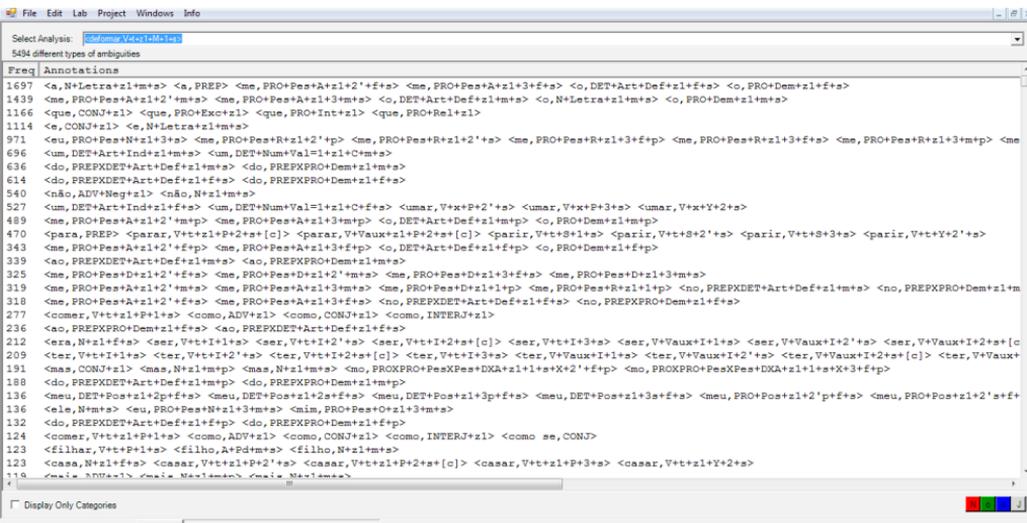
Da leitura das concordâncias relativas à palavra-tema “casa”, como podemos verificar na Figura 1, conclui-se que a forma “casa” corresponde ao nome feminino singular de “casa” e à primeira das três aceções do nome “casa”, patentes no *Dicionário da Língua Portuguesa* da Porto Editora (versão online, <http://www.infopedia.pt/>; “1. qualquer edifício destinado a habitação”). Deste modo, como se pode verificar na Tabela 3, a palavra-tema “casa” insere-se no campo temático de “família”.

# ANÁLISE LEXICOMÉTRICA DE LUANDA, LISBOA, PARAÍSO

**Tabela 3** Campos temáticos de *Luanda, Lisboa, Paraíso*.

Campos temáticos	Palavras-tema
Família	Pai/pai, Casa/casa, filho, Papá, Mãe/mãe, Mulher/mulher, marido
Corporeidade	olhos, mãos, Mão/mão cabeça, corpo, boca, calcanhar, costas, cabelo, dedos, pé, cara
Esperança	dia, menino, criança, miúdo, rapaz, novo, vida, Fazer/fazer, Fazia/fazia, Dizer, Dizia/dizia, saber, Sabia/sabia, Sentia/sentia, ver
Diáspora/Imigração	Vai/vai, la/ia, cidade, lugar, vez, vezes, anos, volta, tempo, lado, meio
Casa	quarto, Cama/cama, porta, mesa
Amizade	amigo, homem, homens, dar, Dava/dava
Solidão	noite, Velho/velho, fim
Desigualdade	Rua/rua, chão, água

Além disso, no Nooj, podemos também visualizar as palavras ambíguas (Figura 2) e as palavras não ambíguas (Figura 3), tendo acesso a todas as anotações produzidas para o *corpus* em estudo.



**Figura 2** Palavras ambíguas de *Luanda, Lisboa, Paraíso*.

Freq	Annotation
1848	<de, PREP>
469	<com, PREP>
298	<em, PREP>
252	<cartola, N+z1+f+a>
192	<pai, N+z1+m+a>
185	<por, PREP>
184	<Aquiles, N+Hum>
139	<sem, PREP>
106	<nem, CONJ+z1>
91	<já, ADV+z1>
86	<num, PREFIXDET+Art+Ind+z1+m+a>
85	<ou, CONJ+z1>
83	<olho, N+z1+m+tp>
82	<dado, DET+Num+Val=2+z1+C+m+tp>
82	<vez, N+z1+f+a>
81	<glória, N+z1+f+a>
80	<ainda, ADV+z1>
71	<dia, N+z1+m+a>
68	<num, PREFIXDET+Art+Ind+z1+f+a>
67	<tempo, N+z1+m+a>
66	<tudo, PRO+Ind+z1>
65	<homem, N+z1+m+a>
62	<ninguém, PRO+Ind+z1>
54	<noite, N+z1+f+a>
54	<depois, ADV+z1>
53	<mão, N+z1+f+tp>
52	<cama, N+z1+f+a>
52	<nunca, ADV+z1>
51	<lugar, N+z1+m+a>
51	<mão, N+z1+f+a>
51	<...>
2833	unambiguous linguistic units

**Figura 3** Palavras não ambíguas de *Luanda, Lisboa, Paraíso*.

As anotações também podem ser visualizadas, palavra a palavra, frase a frase ou parágrafo a parágrafo. Por exemplo, na Figura 4, podemos verificar as anotações da última frase do romance *Luanda, Lisboa, Paraíso*.

## ANÁLISE LEXICOMÉTRICA DE LUANDA, LISBOA, PARAÍSO



Figura 4 Anotações da última frase de *Luanda, Lisboa, Paraíso*.

Tendo por base a referida metodologia, como podemos observar na Tabela 3, no romance *Luanda, Lisboa, Paraíso*, foi possível delimitar oito campos temáticos de diferentes domínios: o campo temático do domínio da “família” foi delimitado através das palavras-tema “Pai”/“pai”, “Casa”/“casa”, “filho”, “Papá”, “Mãe”/“mãe”, “Mulher”/“mulher” e “marido”; o campo temático de “corporeidade” foi delimitado através das palavras-tema “olhos”, “mãos”, “Mão”/“mão”, “cabeça”, “corpo”, “boca”, “calcanhar”, “costas”, “cabelo”, “dedos”, “pé” e “cara”; o campo temático do domínio de “esperança”, a partir das palavras-tema “dia”, “menino”, “criança”, “miúdo”, “rapaz”, “novo”, “vida”, “Fazer”/“fazer”, “Fazia”/“fazia”, “dizer”, “Dizia”/“dizia”, “saber”, “Sabia”/“sabia”, “Sentia”/“sentia” e “ver”; o campo temático de “diáspora/imigração”, relacionado com as palavras-tema “Vai”/“vai”, “la”/“ia”, “cidade”, “lugar”, “vez”, “vezes”, “anos”, “volta”, “tempo”, “lado” e “meio”; o campo temático do domínio de “casa” foi construído a partir das palavras-tema “quarto”, “Cama”/“cama”, “porta” e “mesa”; o campo temático do domínio de amizade foi delimitado através das palavras-tema “amigo”, “homem”, “homens”, “dar” e “Dava”/“dava”; o

campo temático de “solidão” foi delimitado a partir das palavras-tema “noite”, “Velho”/“velho” e “fim”; o campo temático do domínio de “desigualdade” foi construído através das palavras-tema “Rua”/“rua”, “chão” e “água”.

## Conclusões

Neste trabalho, através da lexicometria, identificaram-se as palavras-tema do romance *Luanda, Lisboa, Paraíso* e delimitaram-se os campos temáticos da obra.

Diga-se, por fim, que a ligação entre a lexicometria e o texto literário constitui uma mais-valia para o desenvolvimento e aprimoramento de diversas ferramentas informáticas, de que o Nooj é um exemplo, que subsidiam contributos inegáveis para o progresso das pesquisas. De facto, os procedimentos lexicométricos possibilitam um estudo alternativo e complementar das análises mais tradicionais de cariz hermenêutico. Deste modo, a lexicometria, por constituir uma abordagem que nos permite analisar uma elevada quantidade de dados linguísticos, concede-nos uma percepção distinta da que os pesquisadores geralmente têm ao analisar um texto literário através de métodos não computacionais.

Esperamos que este trabalho possa fornecer um contributo adicional para uma visão mais completa e abrangente do texto literário *Luanda, Lisboa, Paraíso*, mostrando, paralelamente, que os procedimentos lexicométricos configuram um importante aliado no estudo das áreas semânticas e na definição dos temas desta obra literária, em que os campos temáticos de “família”, “corporeidade”, “esperança”, “casa” e “amizade” ocorrem concomitantemente com os campos temáticos de “diáspora/imigração”, “solidão” e “desigualdade”.

## Referências

Almeida, D. P. de. (2018). *Luanda, Lisboa, Paraíso*. Companhia das Letras.

Biber, D., Conrad, S., & Reppen, R. (1998). *Linguistics: Investigating language structure and use*. Cambridge University Press.

## ANÁLISE LEXICOMÉTRICA DE LUANDA, LISBOA, PARAÍSO

- Biderman, M. T. C. (1981). *Teoria linguística: Linguística quantitativa e computacional*. Livros Técnicos e Científicos.
- Carvalho, D., Marques, M. E. R., & Silva, M. F. (1999). Discurso: Práticas lexicométricas. In P. Marrafa & M. A. Mota (Eds.), *Linguística computacional: Investigação fundamental e aplicações* (pp. 255–262). Associação Portuguesa de Linguística; Edições Colibri.
- Galisson, R., & Coste, D. (1983). *Dicionário de didática das línguas* (A. A. Pinto, Trad.). Livraria Almedina. (Trabalho original publicado em 1976)
- Genouvrier, E., & Peytard, J. (1985). *Linguística e ensino do português*. Livraria Almedina.
- Halliday, M. (1973). As bases funcionais da linguagem. In M. Dascal (Ed.), *Fundamentos metodológicos da linguística* (pp. 125–161). Global Universitária.
- Halliday, M. (1985). *An introduction to functional grammar*. Edward Arnold.
- Leblanc, J. M. (2015). Proposition de protocole pour l'analyse des données textuelles: Pour une démarche expérimentale en lexicométrie. *Nouvelles Perspectives en Sciences Sociales*, 11(1), 25–63. <https://doi.org/10.7202/1035932ar>
- Marchand, P. (2013). Quelques traces chronologiques de l'exploration textométrique. *Bulletin de Méthodologie Sociologique*, 120(1), 38–46. <https://doi.org/10.1177/0759106313497856>
- Mateus, M. H. M. (2003). *Gramática da língua portuguesa*. Caminho.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot.
- Niklas-Salminen, A. (1997). *La lexicologie*. Armand Colin.
- Rey-Debove, J. (1998). *La linguistique du signe. Une approche sémiotique du langage*. Armand Colin.
- Salem, A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, 1(2), 5–28.
- Sardinha, T. B. (2004). *Linguística de corpus*. Editora Manole.
- Shaw, H. (1982). *Dicionário de termos literários* (C. dos Reis, Trad.). Publicações Dom Quixote. (Trabalho original publicado em 1976)
- Silberstein, M. (2003). *Nooj manual*. <https://atishs.univ-fcomte.fr/nooj/files/NoojManual.pdf>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Vilela, M. (1994). *Estudos de lexicologia do português*. Almedina.
- Tamba-Mecz, I. (1998). *La sémantique*. Presse Universitaires de France.