

## Exploração de corpus bilingue português brasileiro-inglês através do LancsBox®: possibilidades para uma pesquisa em Tradução / *Exploring a Portuguese-English corpus through LancsBox®: possibilities for research in translation studies*

*João Gabriel Carvalho Marcelino* \*

Doutorando em Estudos da Tradução, Mestre em Linguagem e Ensino (PPGLE/UFCG). Realiza pesquisas no campo dos Estudos da Tradução, com foco em Tradução Literária, Tradução Interlingual no par Português – Inglês, Tradução e ferramentas computacionais. Atualmente, está realizando pesquisa de doutorado sobre a tradução de elementos do sertão nordestino de Vidas Secas para a língua inglesa.

 <https://orcid.org/0000-0001-6528-0208>

**Recebido** em 10 out. 2022. **Aprovado** em: 31 jan. 2023.

### Como citar este artigo:

MARCELINO, João Gabriel Carvalho. Exploração de corpus bilingue português brasileiro-inglês através do LancsBox®: possibilidades para uma pesquisa em Tradução. v. 12, n. 1, *Revista Letras Raras*. Campina Grande, v. 12, n. 1, p. 158-180, abr. 2023.

### RESUMO

Diante da variedade de ferramentas de processamento linguístico, de acesso livre ou pago, neste artigo observa-se a exploração de um corpus de tradução a partir da ferramenta LancsBox®, desenvolvida na Lancaster University. Buscando explorar um corpus paralelo bilingue de pesquisa em tradução utilizando a ferramenta de processamento de linguagem LancsBox®, caracterizam-se aqui os seguintes objetivos específicos: i) Discutir a utilização de aplicativos e ferramentas de processamento linguístico através da Linguística de Corpus e da Linguística Computacional; e ii) Sugerir possibilidades de aplicação do LancsBox® na pesquisa em tradução. Esse estudo experimental está fundamentado em Brezina, McEnery e Wattam (2015), Brezina e McEnery (2021), Berber Sardinha (2002) e Jurafsky e Martin (2022). Os resultados da exploração evidenciam possibilidades de aplicação das funções da ferramenta através de buscas por termos que causam opacidade, utilizando a busca pelas expressões de maneira simples ou através de exploração por RegEx. Os resultados das buscas mostram que ferramentas de processamento linguístico podem enriquecer as análises de traduções, permitindo localizar de maneira mais eficiente as ocorrências buscadas.

**PALAVRAS-CHAVE:** Linguística de Corpus; Vidas Secas; Barren Lives; Estudos da Tradução.

---

\*

 [joaogabrielcarvalho@hotmail.com](mailto:joaogabrielcarvalho@hotmail.com)

 [10.5281/zenodo.7869579](https://doi.org/10.5281/zenodo.7869579)

## ABSTRACT

Given the variety of linguistic processing tools, with free or paying access, in this paper presents the exploration of a translation corpus using the LancsBox® tool developed at Lancaster University. Seeking to explore a bilingual parallel corpus of research in translation using LancsBox® tool; the specific objectives are: i) To discuss the use of applications and linguistic processing tools through Corpus Linguistics and Computational Linguistics; and ii) To point out the possibilities of application of LancsBox® in translation research. This experimental study is based on Brezina, McEnery and Wattam (2015), Brezina and McEnery (2021), Berber-Sardinha (2002) and Jurafsky and Martin (2022). Findings show different application possibilities of the tool's functions through the search for terms that cause opacity, using the search for expressions in a simple way or through exploration by RegEx. Results have shown that linguistic processing tools can enrich the analysis of translations, facilitating the localization of the occurrences more efficiently.

**KEYWORDS:** Corpus Linguistics. Vidas Secas. Barren Lives. Translation Studies.

## 1 Introdução

Os Estudos da Tradução têm se beneficiado do contato com a Linguística Computacional e a Linguística de corpus e, através desse contato encontra-se a possibilidade de utilização de aplicativos de processamento linguístico, como o AntConc®, o WordSmith tools®, entre outras. Diante disto, este artigo utiliza o aplicativo LancsBox®, desenvolvido pela Lancaster University, como modelo para explorar um *corpus* bilíngue de pesquisa em Tradução, buscando apresentar algumas possibilidades de aplicação para pesquisas em Estudos da Tradução. Entretanto, tais aplicativos foram elaborados em sua maioria para a realização de pesquisas com *corpora* monolíngues, cabendo aos pesquisadores dos Estudos da Tradução adaptarem-se para utilizar as ferramentas que compõem tais aplicativos em pesquisas voltadas para *corpora* textuais mono e bilíngues. Observando que cada pesquisa possui particularidades que orientam a escolha de programas mais adequados, diante da variedade de ferramentas que os aplicativos oferecem.

Nesse sentido, este artigo apresenta uma exploração de *corpus* bilíngue no par português-inglês para expor possibilidades de utilização da ferramenta LancsBox® considerando uma pesquisa orientada aos Estudos da Tradução. Utilizando como modelo *para* a exploração o *corpus* bilíngue elaborado a partir da obra Vidas Secas, de Graciliano Ramos, e sua respectiva tradução. Tendo em vista que cada pesquisa observa a língua em um aspecto diferente, este artigo busca contribuir com possibilidades de utilização de ferramentas e estratégias de busca, contemplando *corpora* de pesquisa em tradução e ferramentas computacionais que otimizam as metodologias de pesquisa linguística.

Como objetivo geral, o artigo visa explorar um *corpus* paralelo bilíngue de pesquisa em tradução utilizando a ferramenta de processamento de linguagem LancsBox®. Para tanto, foram definidos os seguintes objetivos específicos: i) Discutir a utilização de aplicativos e ferramentas de

processamento linguístico através da Linguística de *Corpus* e Linguística computacional; e ii) Apontar as possibilidades de aplicação do *LancsBox*® na pesquisa em tradução.

Este artigo fundamenta-se nas pesquisas de Brezina, McEnery e Wattam (2015), Brezina e McEnery (2021), Berber Sardinha (2002), entre outros. Tendo caráter experimental está dividido em quatro seções. A primeira seção apresentando a fundamentação teórica sobre Linguística Computacional e Linguística de *corpus*, bem como a relação entre Linguística de *Corpus* e os Estudos da Tradução. A segunda seção apresenta a Metodologia, a ferramenta *LancsBox*® e o *corpus* estudado. A terceira seção apresenta os resultados das buscas realizadas na ferramenta, e por fim, as considerações finais.

## 2 Linguística Computacional e de Linguística de Corpus

O desenvolvimento de tecnologias computacionais como, o microcomputador, possibilitou o surgimento de novas abordagens a problemas no estudo das línguas. Esse advento permitiu que os *corpora* eletrônicos fossem construídos ou digitalizados, que pesquisas em *corpora* linguísticos fossem realizadas de maneira mais rápida para observar regularidades na língua, assim como permitiu que programas capazes de interpretar e gerar informações sobre linguagem natural fossem desenvolvidos (VIEIRA, LIMA, 2001). Nesse sentido, a linguística computacional tem se envolvido com o desenvolvimento de ferramentas de software voltados ao processamento de linguagem natural, aplicáveis aos diferentes campos de pesquisa linguística.

Por sua vez, a utilização de *corpora* como recurso para a pesquisa linguística tem sido recorrente (ALUÍSIO; ALMEIDA, 2021) tendo em vista as diferentes observações possíveis sobre a linguagem e seu funcionamento, mesmo antes da aparição das ferramentas computacionais. Nesse sentido, McEnery e Hardie (2012) definem a Linguística de *Corpus* como a área focada em desenvolver uma série de procedimentos e métodos para estudar linguagem, tornando possível aos linguistas utilizar tais procedimentos e métodos aplicados a diferentes *corpora* para responder questionamentos de pesquisa aplicados a diversos campos disciplinares alinhados a linguística.

A linguística tem definido os *corpora* com diferentes caracterizações ao longo do tempo, podendo classificá-los como um conjunto de *textos* finito que pode ser tomado como objeto de análise; ou para estabelecer a gramática descritiva de uma língua; ou um conjunto de enunciados emitidos em uma língua; ou até mesmo um conjunto de *textos* escritos ou falados que podem ser

disponibilizados para análise (ALUÍSIO; ALMEIDA, 2021). Tais caracterizações convergem na ideia de que um *corpus* é um conjunto de *textos* que podem ser analisados para diferentes fins, considerando as diferentes ramificações da linguística, como a Tradução no caso deste artigo.

A linguística de *corpus* necessita da relação com outro campo para que as análises sejam, de fato, realizadas, o que fica evidente quando se observa que os procedimentos ainda são considerados como ‘em desenvolvimento’, apesar de alguns já possuírem maiores ou menores graus de consolidação, como o concordanciamento<sup>1</sup> (MCENERY; HARDIE, 2012).

Tais procedimentos são desenvolvidos com o auxílio da Linguística Computacional e Desenvolvimento de Software para aplicações como AntConc, *Wordsmith*, *LancsBox*, entre outras, que realizam o processamento dos *textos* de *corpora* digitalizados. A utilização de *corpora* mostra a inexatidão da intuição humana no entendimento da linguagem, o que pode ser observado com maior precisão através de ferramentas eletrônicas (BERBER SARDINHA, 2002).

Os *corpora* devem seguir algumas características em sua elaboração para a posterior análise. Aluísio e Almeida (2021) apontam como questões importantes para a elaboração de um *corpus*: i) Autenticidade dos *textos*; ii) Representatividade do *Corpus*; iii) Balanceamento (correspondendo ao equilíbrio dos tipos ou gêneros discursivos e *textuais*); iv) Amostragem; v) Diversidade; e vi) Tamanho adequado a pesquisa. Essas características são direcionadas pelo objetivo deste artigo, bem como apontam os métodos de coleta de dados.

As características de *corpora* apontadas por Aluísio e Almeida (2021) corroboram as ideias de Knight (2011), que aponta quatro características dos *corpora* paralelos: i) design e infraestrutura, correspondendo à estrutura do *corpus* desde a construção até a apresentação; ii) tamanho e objetivo, tratando das dimensões do *corpus*; iii) natureza, tratando do quão realista o *corpus* é para o contexto pesquisado; e iv) disponibilidade e (re)utilização, considerando se o *corpus* é publicado na íntegra ou em partes e se está disponível para outros pesquisadores para verificação ou análise. Essas características possibilitam estruturar um *corpus* de pesquisa linguística em diferentes formatos e para diferentes objetivos, apontando a possibilidade da aplicação de diferentes olhares sobre um mesmo *corpus*.

Aluísio e Almeida (2021) consideram três etapas para a coleta de um *corpus* i) projeto do *corpus*: ii) compilação, manipulação, nomeação dos arquivos e pedidos de permissão de uso; e iii) anotação estrutural ou linguística do *corpus* de trabalho. Essas etapas se consolidam na

---

<sup>1</sup> Concordanciamento refere-se ao processo de verificar como palavras se comportam em um *texto*, realizado por ferramentas de processamento linguístico como o *LancsBox*, *AntConc*, entre outras. Os concordanciadores permitem observar padrões sintáticos ou de colocação de palavras para diferentes tipos de análise linguística.

construção de um *corpus* de pesquisa linguística que pode ser tratado e processado através de ferramentas de Processamento de Linguagem Natural (PLN) (ALUÍSIO; ALMEIDA, 2021), alinhando a Linguística de *Corpus* com a Linguística Computacional. Diante do exposto, a seção a seguir explora a relação entre os Estudos da Tradução e a Linguística de *Corpus*, considerando as possibilidades de aplicação de ferramentas computacionais para o campo e a proposta apresentada neste artigo.

## 2.1 Estudos da Tradução e Linguística de *Corpus*

Os Estudos da Tradução, de acordo com Williams e Chesterman (2010), descrevem o campo de disciplinar dedicado à descrição, análise e teorização dos processos, contextos e produtos da tradução, bem como sobre os papéis dos agentes envolvidos nesses espaços de observação. Tal caracterização do campo disciplinar possibilita que diferentes teorias sejam aplicadas sobre os objetos dos Estudos da Tradução, permitindo observar o processo, o produto e o agente da tradução. Dentre essas teorias, a Linguística de *Corpus* e a Linguística Computacional podem ser aplicadas para dar suporte às análises sobre o *texto* traduzido em relação ao *texto* de partida; em relação aos *textos* produzidos na língua de chegada ou de partida, bem como no desenvolvimento de ferramentas de tradução e treinamento e formação de tradutores.

Berber Sardinha (2002) reforça que, apesar de ser uma ferramenta bastante útil, a integração entre os Estudos de *Corpora* e os Estudos da Tradução foi lenta. Nesse sentido, o autor sugere três hipóteses para justificar essa lenta integração. A primeira hipótese sendo o preconceito dos linguistas de *corpus* em relação ao *texto* traduzido, pois os consideravam um *texto* desviante e não representativo da linguagem; a segunda hipótese estando associada à imagem negativa que a Linguística possuía para os tradutores e pesquisadores de tradução, tendo em vista que por muito tempo a Tradução foi vista como uma mera aplicação de diferentes teorias linguísticas; e a terceira hipótese estava na esfera do acesso à tecnologia.

Berber Sardinha (2002) cita os direcionamentos de pesquisa sobre a aplicação dos Estudos de *Corpora* na tradução em tópicos, como a elaboração de *corpora* bi- ou multilíngues; alinhamento automático de *corpora* paralelos; criação e utilização de concordanciadores e outras ferramentas de processamento de *corpora*; Tradução Automática e Memórias de Tradução;

*Corpora* na formação de tradutores e pesquisadores de tradução; e a observação do processo de tradução de tradutores profissionais.

Olohan (2004) enfatiza que tais direcionamentos devem ser bem definidos para a pesquisa, uma vez que os elementos observados em uma pesquisa sobre tradução devem ser claramente delimitados e o *corpus* deve ser construído objetivando uma grande representatividade para os aspectos estudados. Nesse sentido, a definição do tipo de abordagem de pesquisa em Tradução condiciona o método de coleta e anotação do *corpus*; definição de formato de saída; bem como as ferramentas empregadas, sejam para a realização de traduções ou para a análise de traduções.

As possibilidades de métodos de aplicação de *corpora* na pesquisa em tradução são variadas. Berber Sardinha (2002) sugere: i) Identificação da padronização formal de L1 e das funções correspondentes; ii) Identificação de equivalentes tradutórios *prima facie* para cada função; e iii) Identificação da padronização formal de L2 e das funções correspondentes. Outra metodologia apresentada pelo autor referenciando Tognini-Bonelli (2001, 2002)<sup>2</sup> é a baseada em três *corpora*: i) um paralelo, com *textos* na língua de origem e suas respectivas traduções na língua-alvo; ii) um de referência, com *textos* produzidos na língua de origem; e iii) um de referência, com *textos* produzidos na língua de chegada. Essa metodologia sugere que sejam mapeados os padrões relativos aos itens de interesse nos componentes do *corpus* paralelo, bem como dos *corpora* de referência para identificar se os usos presentes no *corpus* paralelo são típicos ou não.

## 2.2 Possibilidades de buscas em uma Tradução Literária

As ferramentas desenvolvidas e exploradas pela Linguística computacional auxiliam na otimização de pesquisas em Linguística Aplicada, como observado na tradução, permitindo que, através dessas ferramentas, pesquisadores encontrem padrões recorrentes na tradução para compreender escolhas no processo tradutório e teorizar sobre os impactos no produto da tradução. Nesse sentido, explorar uma tradução como a que se propõe neste artigo requer posicionar o *texto* e os elementos que guiam as observações realizadas.

---

<sup>2</sup> O autor referencia Elena Tognini-Bonelli com base no livro *Corpus Linguistics at Work*, publicado em 2001 pela editora John Benjamins; e no capítulo de livro *Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach*, parte do livro *Lexis in Contrast: Corpus-Based Approaches*, organizado por in Bengt Altenberg e Sylviane Granger, e publicado pela editora John Benjamins em 2002.

A obra *Vidas Secas* (1938), de Graciliano Ramos, apresenta na narrativa episódica a vida da família de Fabiano e Sinhá Vitória, retirantes que viajam pela caatinga acompanhados dos dois filhos e da cachorra Baleia. A obra de Graciliano Ramos apresenta na narrativa relativamente breve a ambientação da caatinga, bioma exclusivamente brasileiro, que dita os caminhos percorridos pelos personagens. A narração em terceira pessoa e a pouca habilidade das personagens em se expressar tornam o *texto* breve e sucinto, escrito por um “escritor que só dizia o essencial e, quanto ao resto, preferia o silêncio” (CÂNDIDO, 2012, p. 142). Deixando nos breves espaços da escrita elementos tão particulares ao sertão que, na tradução, o tradutor precisa lidar com essa brevidade de palavras para transpor os sentidos do *texto* literário.

A tradução explorada neste artigo foi realizada em 1964 por Ralph Edward Dimmick e apresenta elementos que possibilitam discutir as estratégias adotadas no processo tradutório. Justificando, assim, a utilização de ferramentas computacionais visando explorar o *corpus* para identificar diferentes estratégias de tradução da língua portuguesa para o inglês.

Oustinoff (2011) expõe que não existe uma tradução ‘neutra’ ou ‘transparente’, em que o *texto* original apareceria espelhado, de maneira idêntica. Nesse sentido, a tradução lida com a complexidade dos sentidos na transposição entre línguas com diferentes culturas, ainda mais, na tradução literária, o tradutor lida com as diferentes interpretações que o *texto* permite a cada leitor, uma vez que o *texto* literário não é estático nas interpretações (BRITTO, 2016).

Mesmo em um *texto* de um autor como Graciliano Ramos que buscava dizer somente o essencial, a presença de elementos que não possuem correspondência na língua e cultura de chegada denotam a complexidade da língua e cultura fonte, o que evidencia a natureza criativa do trabalho da tradução (BRITTO, 2016). Para lidar com a seca do ambiente, do *texto* e do bioma, o tradutor precisa realizar escolhas e adotar estratégias que ora apagam o estrangeiro, ora o trazem à tona.

Observando o processo tradutório no par português-inglês, é possível refletir sobre o reconhecimento das relações assimétricas em projetos de tradução, considerando a hegemonia global do inglês (VENUTTI, 2019). Isso permite que as traduções de *textos* da língua portuguesa que fazem parte do sistema literário brasileiro e foram traduzidos para a língua inglesa sejam estudados visando observar como o tradutor lida com elementos particulares da língua e cultura fonte na transposição para a língua e cultura alvo. O que pode evidenciar a tendência de tradução que anexa o *texto* a uma cultura de chegada submetendo-o a essa língua e cultura, o que Berman



(2013) expõe como a ideia de traduzir de maneira a dar impressão de que o autor teria escrito daquela forma, se escrevesse para aquela língua.

Isso permite refletir sobre os processos de Domesticação e Estrangeirização explorados por Venuti (2021) que, respectivamente, reduz o *texto* estrangeiro aos valores da língua alvo incorporando o autor a aquele ambiente; ou cria uma pressão naqueles valores estrangeiros para registrar a diferença linguística e cultural do *texto* estrangeiro, levando o leitor para fora. Observar a presença ou o apagamento do estrangeiro no *texto* permite denotar que abordagem predomina na tradução, permitindo ao pesquisador identificar e demarcar no *corpus* ocorrências que relatem essas tendências para quantificá-las e analisá-las. Portanto, considerar esses dois valores em discussão sobre o *texto* literário, não significa que eles não sejam comportados em algum nível na tradução (BERMAN, 2013), mas que em determinados momentos estratégias distintas podem ser utilizadas para realizar uma tradução funcionalmente equivalente.

É possível localizar em um *corpus* de tradução elementos que ocorrem no *texto* de partida, demarcá-los com ferramentas que permitem a anotação de *corpora* para identificar tendências deformadoras, que Berman (2013) expõe como um jogo de forças inevitável que o tradutor está no meio para transpor a letra (o *texto*), considerando que as línguas na tradução operam de maneira diferente. Bem como, observar o que Franco-Aixelá (2013) discute como Itens Culturais-específicos, elementos que causam opacidade ideológica ou cultural que requerem estratégias diferentes para lidar com eles no processo tradutório.

Nesse sentido, ao reconhecer o tipo de estratégia de tradução ou tendência utilizada, o pesquisador pode reconhecer padrões iniciais para realizar buscas utilizando ferramentas como o a busca por expressões regulares (RegEx) para realizar anotações, encontrar frequências ou padrões no *corpus* a partir de palavras, construções ou partes de palavras que podem ser visivelmente frequentes no processo tradutório. Considerando, por exemplo, a repetição de itens culturais-específicos através da estratégia de conservação ou de explicação intratextual (FRANCO-AIXELÁ, 2013) ou refletir sobre analítica da tradução e tendências deformadoras (BERMAN, 2013) na tradução ou conservação de itens culturais-específicos como os elementos do sertão nordestino presentes em *Vidas Secas* na tradução para *Barren Lives*.

Na seção a seguir, estão apresentados os aspectos metodológicos do artigo, explicitando a ferramenta utilizada, o LancsBox®, o *corpus*, as definições das buscas realizadas para a discussão dos resultados.



### 3 Metodologia

Esta seção apresenta os aspectos metodológicos do artigo, desse modo em cada subdivisão estão apresentados os elementos utilizados para a realização das buscas que são apresentadas na seção de Resultados e Discussões. Desse modo, esta metodologia apresenta o aplicativo LancsBox®, com a descrição das funções e da utilização; e apresentação da interface (BREZINA; MCENERY; WATTAM, 2015; BREZINA; TIMPERLEY; MCENERY, 2018; BREZINA; WEILL-TESSIER; MCENERY, 2021). Em seguida o *corpus* de trabalho é apresentado, descrevendo as etapas de elaboração e os dados estatísticos que descrevem o *corpus*; e, por fim, a lógica de busca é descrita explicitando os itens buscados no *corpus* utilizando a localização de expressões regulares (RegEx).

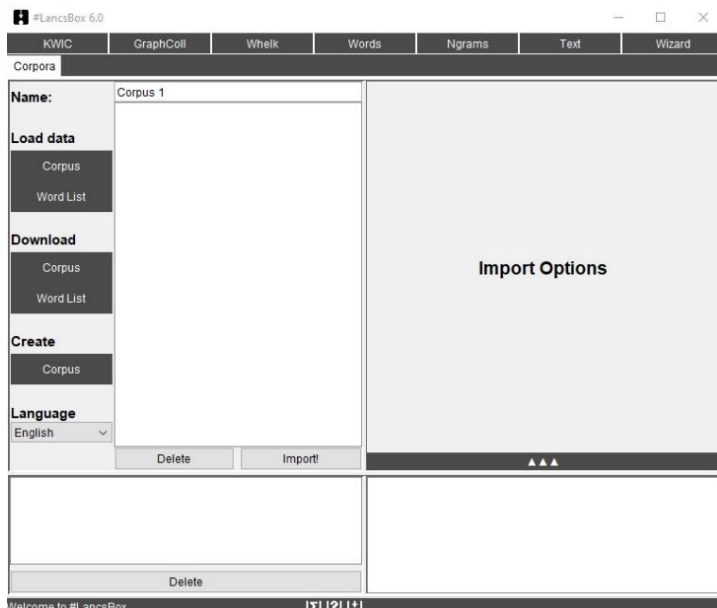
#### 3.1 O aplicativo LancsBox®

*LancsBox*, contração referente à *Lancaster University Corpus Toolbox*, é uma ferramenta de análise de *corpus* desenvolvida pela *Lancaster University*, para sistemas operacionais de 64-bit. A ferramenta é capaz de analisar *corpora* linguísticos em diferentes formatos (.txt, .xml, .doc, .docx, .pdf, .odt, .xls, .xlsx etc.), bem como elaborar *wordlists* e *corpora* extraídos de páginas da internet (BREZINA, 2021)<sup>3</sup>.

A interface do aplicativo *LancsBox*® apresenta, na tela inicial, a opção de adicionar *corpora* de trabalho, bem como as ferramentas disponibilizadas pelo aplicativo e a opção de modificar a linguagem do aplicativo; entretanto a língua portuguesa ainda não está completamente disponível. Por essa razão, o aplicativo foi executado em inglês. Na esquerda são apresentadas as ferramentas de inclusão de *corpus* de trabalho, sendo possível adicionar um *corpus* já existente, fazer o download de um *corpus* diretamente no aplicativo e a criação de *corpus* diretamente no aplicativo. Os *corpora* inseridos no aplicativo ficam salvos e listados na área inferior esquerda, possibilitando acessar novamente quantas vezes forem necessárias:

Figura 1 – Interface do *LancsBox*® 6.0

<sup>3</sup> Disponível em: <http://corpora.lancs.ac.uk/lancsbox/>



Fonte: captura realizada pelo autor

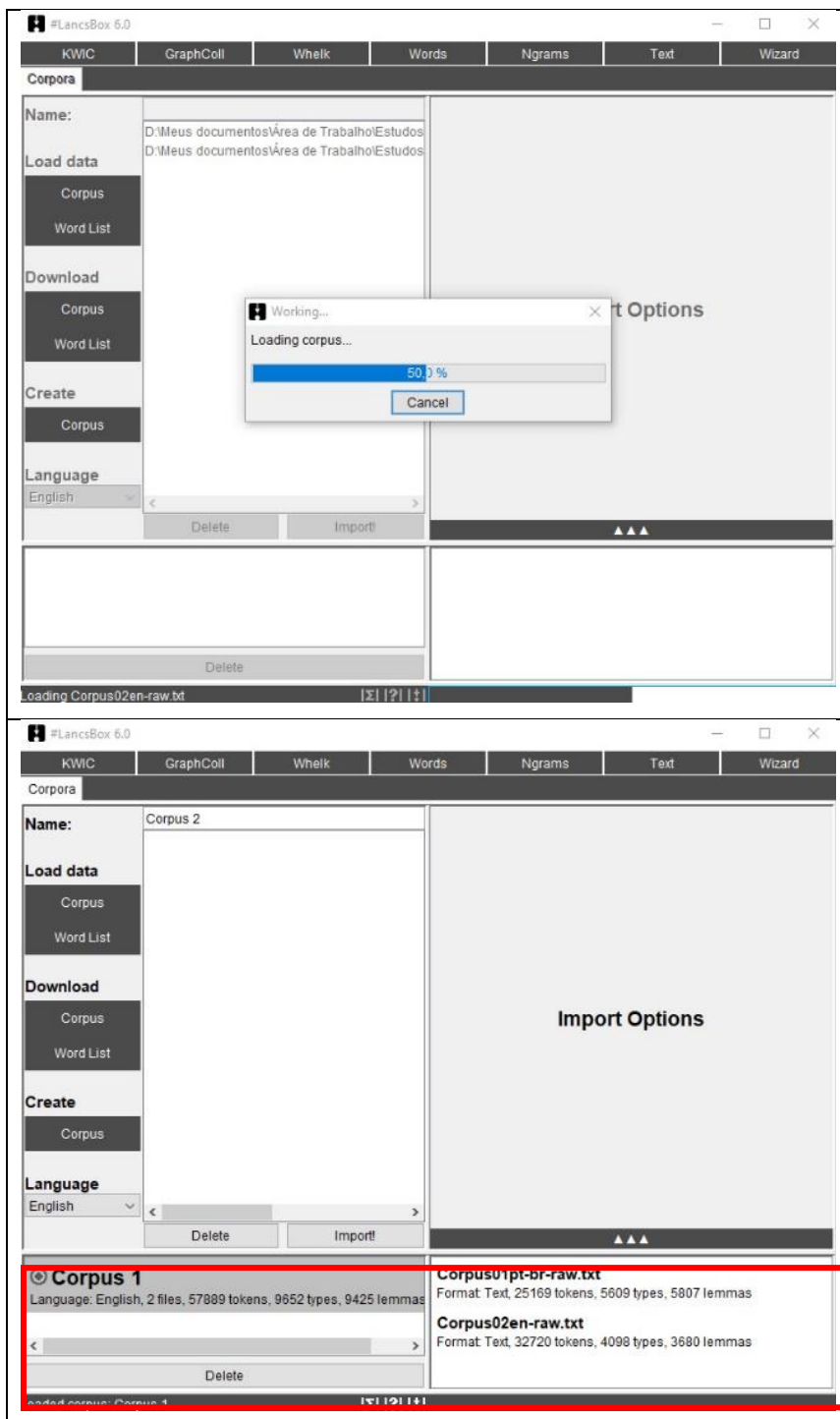
As ferramentas de busca disponíveis na barra superior da interface apresentada na Figura 1 direcionam o usuário para as seguintes opções:

- **KWIC**: busca por palavras-chave em concordâncias
- **GraphColl**: busca colocações e apresenta os resultados em gráfico ou network.
- **Whelk**: mostra a distribuição do termo buscado nos arquivos do *corpus*.
- **Words**: permite observar frequências e realizar a comparação entre *corpora* em combinação com outras ferramentas.
- **Ngrams**: Permite buscar e analisar frequências de *Ngrams*.
- **Text**: os termos da busca são apresentados em contextos frasais.
- **Wizard**: possibilita realizar todas as buscas anteriores e criar um relatório automático com os resultados obtidos. (BREZINA, 2021)

Para a utilização da ferramenta, o *corpus* de trabalho foi adicionado através da opção *Load data >> Corpus* o *corpus* bilíngue utilizado, por se tratar de um *corpus* relativamente pequeno. O

upload no aplicativo foi rápido e foram criados três *corpora* de trabalho, um com o *texto* em português-brasileiro, um em inglês e um com os dois *corpora*:

Figura 2 – Processo de upload de *corpus* no aplicativo



Fonte: elaborado pelo autor

Ao realizar o *upload* dos arquivos de um *corpus*, o *LancsBox* automaticamente realiza a contagem de *token*, *type* e *lemma*<sup>4</sup>, apresentado no quadro com o destaque em vermelho. Os dados apresentados são utilizados para a construção de análise sobre o *corpus* em diferentes etapas de análise. Na seção a seguir, serão apresentados os dados do *corpus* utilizado na exploração proposta aqui apresentada.

### 3.2 O Corpus *Vidas Secas-Barren Lives*

O *corpus* bilíngue elaborado de *Vidas Secas-Barren Lives* é um *corpus* elaborado para a realização de uma pesquisa no campo dos Estudos da Tradução, e corresponde ao *texto* da obra *Vidas Secas* (1938) de Graciliano Ramos, em português-brasileiro e, em paralelo, o *texto* da obra *Barren Lives* (1964), tradução de *Vidas Secas* realizada por Ralph Edward Dimmick. O *corpus* é elaborado a partir de dados que já passaram pela etapa de coleta de dados linguísticos (digitalização, limpeza e pré-processamento), estatística de *corpus* (caracterização quantitativa) e etiquetagem (manual ou automatizada), o que o torna um *corpus* pronto para a etapa de análise linguística (CUNHA, 2020). Na tabela 1, estão descritas as etapas e atividades realizadas na construção do *corpus* de maneira simplificada:

Tabela 1 – Etapas de elaboração do *corpus*

<b>Dados linguísticos (digitalização, limpeza e pré-processamento)</b>	Conversão dos arquivos nato-digitais para .pdf e .txt. Extração do <i>texto</i> e separação dos capítulos. Limpeza do <i>texto</i> . Elaboração do <i>corpus</i> para anotação. Elaboração do <i>corpus</i> sem anotação.
<b>Estatística de <i>corpus</i> (caracterização quantitativa)</b>	Tamanho em kb. Quantidade de palavras. Quantidade de <i>Lemmas</i> , <i>Types</i> e <i>Tokens</i> . Quantidade de linhas.
<b>Etiquetagem (manual ou automatizada)</b>	Anotação de etiquetas.
<b>Análise linguística</b>	Análise voltada para os estudos da tradução.

Fonte: elaborada pelo autor a partir de Cunha (2020) e dos dados do *corpus*.

<sup>4</sup> Token corresponde a “palavras” em um *corpus*, é utilizado para contar palavras no *corpus*; *Type* corresponde a formas particulares de palavras em um *corpus*; e *Lemma* a derivações de palavras em um *corpus* (pedrinha e pedreira são *lemmata* de pedra.).

Para a construção do *corpus textual* foi utilizada a ferramenta *Sublime text*<sup>5</sup>, ferramenta que possibilita escrever *textos* com a codificação UTF-8 (formato universalizado), realizar buscas por expressões regulares, bem como converter para diferentes formatos, incluindo XML (eXtensible Markup Language), formato utilizado para realizar a etiquetagem com *tags*. Diante das etapas apresentadas por Cunha (2020), o *corpus* elaborado apresenta os seguintes dados:

**Tabela 2** – Dados do *corpus Vidas Secas-Barren Lives*

Língua	Português-brasileiro	Inglês
Nome	<i>Corpus01pt-br</i>	<i>Corpus02en</i>
Formato	.txt	.txt
Codificação	UTF-8	UTF-8
Tamanho	164kb	193kb
Palavras	25290	32783
Linhas	2077	2115
Tokens	25169	32720
Types	5609	4098
Lemmas	5807	3680

Fonte: Elaborado pelo autor

Diante dos dados apresentados na Tabela 2, as buscas serão realizadas considerando ambos os *corpora* para observar como a tradução lida com expressões e palavras que causam opacidade no processo de tradução entre a língua de partida e a língua de chegada. Na seção a seguir estão explicitados os termos e as justificativas para a realização das buscas.

### 3.3 Buscas realizadas

Tendo em vista que o *LancsBox*<sup>®</sup> possibilita a realização de buscas utilizando *RegEx* (Expressões regulares), as buscas realizadas aqui são baseadas, também, em expressões regulares. *RegEx*, em síntese, corresponde a buscas avançadas realizadas a partir de qualquer combinação de caracteres apresentadas entre barras (//) utilizada para localizar a concatenação

<sup>5</sup> O *Sublime text* é um editor de código fonte multiplataforma que suporta diferentes linguagens de programação e diferentes linguagens de marcação como o XML. O aplicativo está disponível para download e compra de licença no website do desenvolvedor <https://www.sublimetext.com/>. A versão gratuita do aplicativo não atrapalha a utilização em relação à versão paga.

de caracteres buscados (BREZINA; WEILL-TESSIER; MCENERY, 2020; JURAFSKY, MARTIN; 2022).

As buscas por RegEx, no contexto da Linguística de *Corpus*, permitem especificar elementos linguísticos no *texto* para extraí-los (JURAFSKY, MARTIN; 2022), facilitando a localização e quantificação de ocorrências no *corpus* estudado. Nesse sentido, o que é determinado como expressões regulares para este contexto baseia-se em partículas que demarcam decisões no processo tradutório, como conservação na tradução (localizados através de termos identificados analogicamente, ou radicais, prefixos e sufixos localizados mecanicamente) ou expressões que complementem sentido de palavras conservadas, estabelecendo uma relação de explicação. Estabelecem-se, assim, as seguintes categorias de busca para a exploração apresentada a seguir:

- i) conservação;
- ii) prefixação e sufixação;
- iii) colocação ou ocorrências frequentes.

As buscas apresentadas na seção a seguir estão dispostas de acordo com o critério e a expressão regular buscada, apresentando as imagens resultantes das buscas realizadas na ferramenta *LancsBox*.

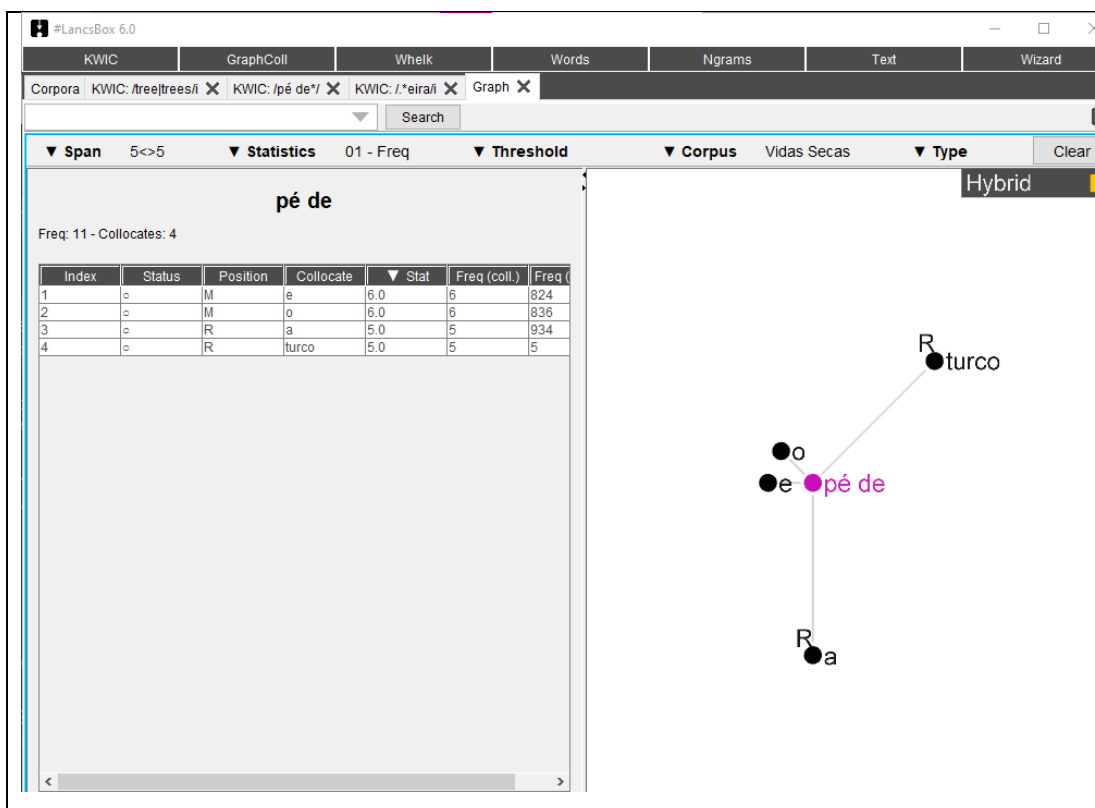
## 4 Resultados e Discussão

Diante do exposto nas seções anteriores, a seguir estão apresentadas as buscas e resultados obtidos, os dados apresentados seguem a seguinte ordem: apresentação do termo ou expressão regular buscado; explanação da lógica da busca; justificativa para a realização da busca, resultado no aplicativo e resultados comentados. As ferramentas utilizadas foram o KWIC (*Keyword in context*) e *GraphColl*, para demonstração.

### 4.1 *GraphColl*

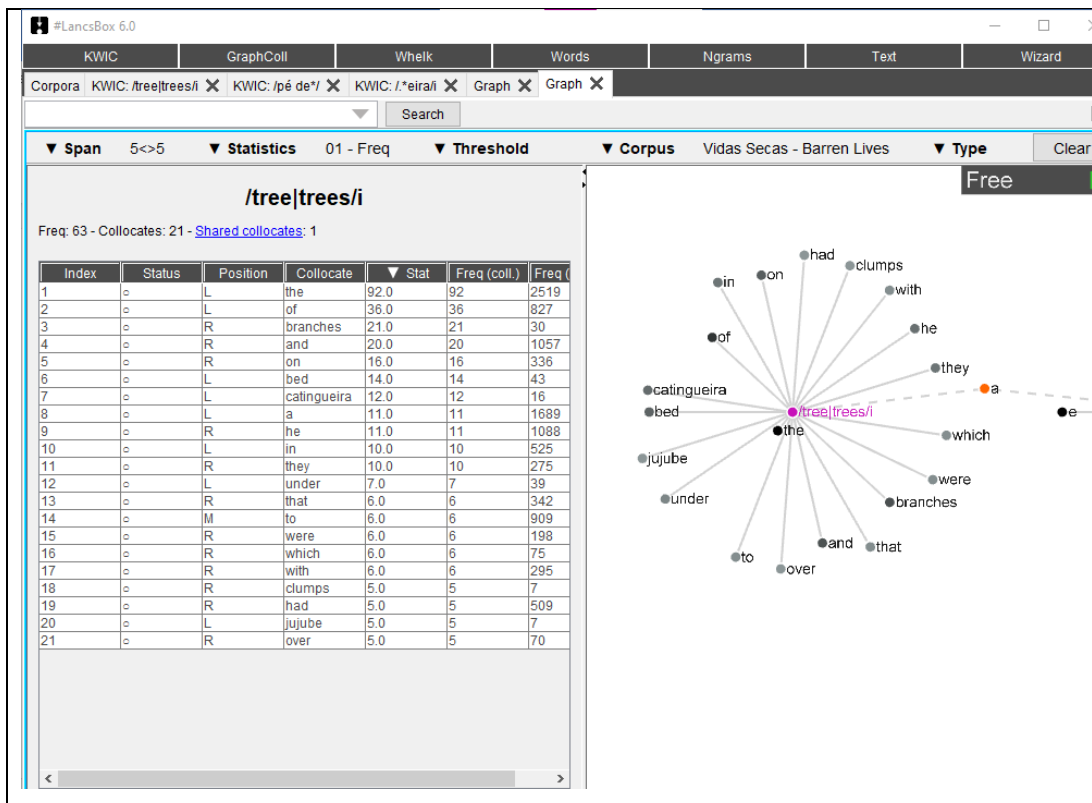
As buscas utilizando a ferramenta *GraphColl* foram realizadas no *corpus* bilingue para localizar colocações recorrentes com as palavras ‘pé de’ presentes no *corpus* em português brasileiro, e ‘tree’/‘trees’ presentes no *corpus* em inglês. As expressões regulares /pé de/ e /tree|trees/ são formuladas com os termos dispostos na íntegra, considerando a ocorrência separada por barra (|) para considerar o singular e o plural<sup>6</sup>. A busca objetivou localizar construções iniciadas com ‘pé de’, como ‘pé de turco’ ou finalizadas com ‘tree’ ou ‘trees’ para observar a construção de nomes de elementos da vegetação:

Figura 3 – Resultados das buscas na ferramenta *GraphColl* /pé de/ e /tree|trees/



<sup>6</sup> Tendo em vista que foi buscada a expressão tree como complemento a termos em português conservados no *texto* de chegada, não foi necessário considerar a possibilidade de diferenciar maiúsculas de minúsculas, quando necessário, deve-se usar os colchetes como no exemplo: /[t]ree|[t]rees/ (JURAFSKY, MARTIN; 2022).





Fonte: elaborado pelo autor

Os resultados obtidos mostraram a maior ocorrência de construções finalizadas com a expressão *tree* no *texto* traduzido para justificar a conservação de um termo da língua de chegada, realizando uma explicação intratextual (FRANCO-AIXELÁ, 2013) do termo conservado. A menor ocorrência de colocações da expressão “pé de” na parte do *corpus* na língua de chegada evidencia que a construção das palavras por sufixação, terminada em *-eira* naturalmente aponta para elementos da vegetação, tornando não obrigatória a presença da expressão. Nos dados obtidos do *corpus* na língua de chegada, a ocorrência de explicação intratextual permite ao leitor compreender o sentido das palavras conservadas por *contexto*, como no caso da *Catingueira* que em *contexto* de língua portuguesa brasileira, compreende-se como árvore nativa do bioma caatinga, onde se passa a narrativa de *Vidas Secas*.

#### 4.2 Prefixação e Sufixação

As buscas por sufixação realizadas no *texto* de chegada, *Barren Lives*, foram feitas considerando nomes de elementos da vegetação terminados em *'-eira'*. Para tanto, a expressão regular buscada foi */\*eira/* e */\*eiro/*, a utilização do asterisco (\*) para a construção da expressão

regular serve para indicar a ocorrência de zero ou mais ocorrências da sequência de caracteres que ocorre após o asterisco (JURAFSKY, MARTIN; 2022). A busca foi realizada visando identificar a ocorrência de conservação de nomes vernaculares de elementos da vegetação no português brasileiro presentes no *corpus* em inglês:

Figura 4 – retorno das buscas pelos sufixos -eira e -eiro na ferramenta KWIC no *corpus* em inglês



Fonte: Elaborado pelo autor

Quadro 1 – Resultados extraídos do LancsBox®

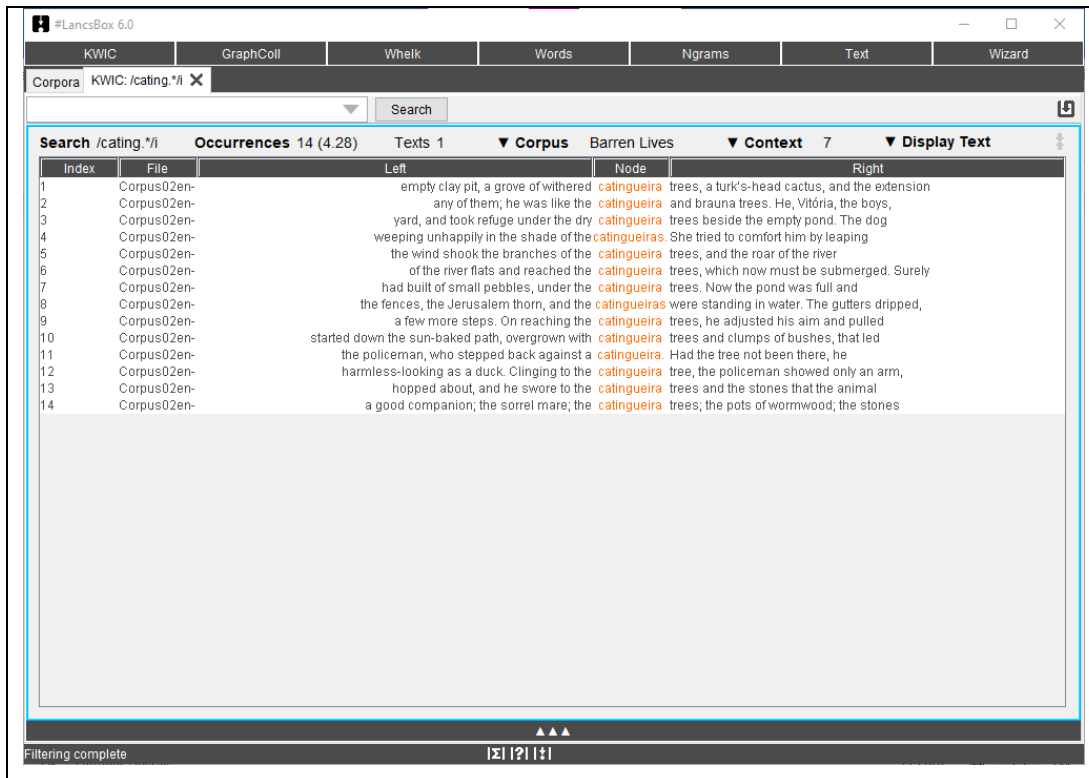
Corpus: <i>Barren Lives</i>   Search Term: <i>/.*eira/i</i>   Occurrences: 12 (3.67)   Texts: 1				
Index	File	Left	Node	Right
1	<i>Corpus02en.txt</i>	empty clay pit, a grove of withered	catingueira	trees, a turk's-head cactus, and the extension
2	<i>Corpus02en.txt</i>	any of them; he was like the	catingueira	and brauna trees. He, Vitória, the boys,
3	<i>Corpus02en.txt</i>	yard, and took refuge under the dry	catingueira	trees beside the empty pond. The dog
4	<i>Corpus02en.txt</i>	the wind shook the branches of the	catingueira	trees, and the roar of the river
5	<i>Corpus02en.txt</i>	of the river flats and reached the	catingueira	trees, which now must be submerged. Surely
6	<i>Corpus02en.txt</i>	had built of small pebbles, under the	catingueira	trees. Now the pond was full and
7	<i>Corpus02en.txt</i>	a few more steps. On reaching the	catingueira	trees, he adjusted his aim and pulled
8	<i>Corpus02en.txt</i>	started down the sun-baked path, overgrown with	catingueira	trees and clumps of bushes, that led
9	<i>Corpus02en.txt</i>	the policeman, who stepped back against a	catingueira.	Had the tree not been there, he
10	<i>Corpus02en.txt</i>	harmless-looking as a duck. Clinging to the	catingueira	tree, the policeman showed only an arm,
11	<i>Corpus02en.txt</i>	hopped about, and he swore to the	catingueira	trees and the stones that the animal
12	<i>Corpus02en.txt</i>	a good companion; the sorrel mare; the	catingueira	trees; the pots of wormwood; the stones

Fonte: elaborado pelo autor

Os resultados obtidos, apresentados na Figura 4 e Tabela 3, apontaram para a conservação da palavra *catingueira*, com a complementação *tree* explicitada na seção anterior. Bem como para a tradução de elementos da vegetação terminados em *-eiro*, como Juazeiro (traduzido na obra como *Jujube tree*), o que na busca realizada no *corpus* da língua de chegada resultou em zero ocorrências. Tendo em vista que a palavra *Catingueira* deriva de *Caatinga*, grafada *Catinga* em *Vidas Secas*, foi realizada a busca por prefixação para identificar a ocorrência da palavra *Catinga*, buscando contrastar os resultados da busca anterior pelo sufixo, utilizando a expressão regular */cating\*/* para o prefixo.

A utilização do asterisco no final do prefixo buscado segue lógica semelhante à da busca por sufixo, estabelecendo a busca baseada em zero ou mais ocorrências da sequência anterior (JURAFSKY, MARTIN; 2022), resultando, portanto, em:

Figura 5 – retorno das buscas pelo prefixo cating- na ferramenta KWIC



Fonte: elaborado pelo autor

Quadro 2 – Resultados obtidos no LancsBox®

Corpus: Barren Lives   Search Term: /cating.*  Occurrences: 14 (4.28)   Texts: 1				
Index	File	Left	Node	Right
1	Corpus02en.txt	empty clay pit, a grove of withered	catingueira	trees, a turk's-head cactus, and the extension
2	Corpus02en.txt	any of them; he was like the	catingueira	and brauna trees. He, Vitória, the boys,
3	Corpus02en.txt	yard, and took refuge under the dry	catingueira	trees beside the empty pond. The dog
4	Corpus02en.txt	weeping unhappily in the shade of the	catingueiras.	She tried to comfort him by leaping
5	Corpus02en.txt	the wind shook the branches of the	catingueira	trees, and the roar of the river
6	Corpus02en.txt	of the river flats and reached the	catingueira	trees, which now must be submerged. Surely
7	Corpus02en.txt	had built of small pebbles, under the	catingueira	trees. Now the pond was full and
8	Corpus02en.txt	the fences, the Jerusalem thorn, and the	catingueiras	were standing in water. The gutters dripped,
9	Corpus02en.txt	a few more steps. On reaching the	catingueira	trees, he adjusted his aim and pulled
10	Corpus02en.txt	started down the sun-baked path, overgrown with	catingueira	trees and clumps of bushes, that led
11	Corpus02en.txt	the policeman, who stepped back against a	catingueira.	Had the tree not been there, he
12	Corpus02en.txt	harmless-looking as a duck. Clinging to the	catingueira	tree, the policeman showed only an arm,
13	Corpus02en.txt	hopped about, and he swore to the	catingueira	trees and the stones that the animal
14	Corpus02en.txt	a good companion; the sorrel mare; the	catingueira	trees; the pots of wormwood; the stones

Fonte: Elaborada pelo autor

As buscas pelo radical Cating- retornaram o mesmo resultado que as buscas por -eira, apresentando mais duas ocorrências devido ao plural. Com os resultados obtidos é possível observar que apesar da conservação do nome da planta seguido da explicação intratextual (FRANCO-AIXELÁ, 2013) para estabelecer a ideia de que ‘Catingueira’ corresponde a um elemento da vegetação, o apagamento do nome do bioma resulta na destruição de uma rede de significados adjacente (BERMAN, 2013), destruindo a relação entre o nome da árvore e o nome do bioma do qual deriva. Ao realizar as mesmas buscas no *corpus* da língua de chegada, a ferramenta KWIC obteve os resultados apresentados na figura 6:

Figura 6 - retorno das buscas pelos sufixos -eira e -eiro na ferramenta KWIC no *corpus* em português

The figure consists of two screenshots of the KWIC search tool interface. The top screenshot shows search results for the suffix '-eiro' with 122 occurrences. The bottom screenshot shows search results for the suffix '-eira' with 137 occurrences. Both screenshots display a table with columns for Index, File, Left, Node, and Right, showing text snippets with the search term highlighted in orange.

Index	File	Left	Node	Right
1	Corpus01-pt-t	verdes. Os infelizes tinham caminhado o dia	riteiro,	estavam cansados e famintos. Ordinariamente andavam pouco,
2	Corpus01-pt-t	culpado, mas dificultava a marcha, e o	vaqueiro	precisava chegar, não sabia onde. Tinham deixado
3	Corpus01-pt-t	mais arrastada, num silêncio grande. Ausente do	ompanheir	a cachorra Baleia tomou a frente do
4	Corpus01-pt-t	fazenda sem vida O curral deserto, o	chiqueiro	das cabras arruinado e também deserto, a
5	Corpus01-pt-t	arruinado e também deserto, a casa do	vaqueiro	fechada, tudo anunciava abandono. Certamente o gado
6	Corpus01-pt-t	plantas mortas, rodeou a tapera, alcançou o	terreiro	do fundo, viu um barreiro vazio, um
7	Corpus01-pt-t	alcançou o terreno do fundo, viu um	barreiro	vazio, um bosque de catingueiras murchas, um
8	Corpus01-pt-t	quis acordá-los. Foi apanhar gravetos, trouxe do	chiqueiro	das cabras uma braçada de madeira meio
9	Corpus01-pt-t	aproveitou os gravetos, arrastou-se sobre o	chiqueiro	da praça, fez um minuto, localizou os

Index	File	Left	Node	Right
1	Corpus01-pt-t	correia presa ao cinturão, a espingarda de	pederneirano	ombro. O menino mais velho e
2	Corpus01-pt-t	areia do rio, onde haviam descansado, a	beira	de uma poça: a fome apertara demais
3	Corpus01-pt-t	aligeirou o passo, esqueceu a fome, a	canseira	e os ferimentos. As alpercatas dele estavam
4	Corpus01-pt-t	do rio, acompanharam a cerca, subiram uma	ladeira,	chegaram aos juazeiros. Fazia tempo que não
5	Corpus01-pt-t	do chiqueiro das cabras uma braçada de	madeira	meio roída pelo cupim, arrancou touceiras de
6	Corpus01-pt-t	touceiras de macambira, arrumou tudo para a	fogueira.	Nesse ponto Baleia arrebitou as orelhas, arregaçou
7	Corpus01-pt-t	couro. Fabiano tomou a cuia, desceu a	ladeira,	encaminhou-se ao rio seco, achou no bebedouro
8	Corpus01-pt-t	bem dizer não se diferenciava muito da	bolandeira	de seu Tomás. Agora, deitado, apertava a
9	Corpus01-pt-t	os dentes. Que fim teria levado a	bolandeira	de seu Tomás? Olhou o céu de
10	Corpus01-pt-t	Tomás fugiu também, com a casa,	bolandeira	estava parada. E ele, Fabiano, arrastou

Fonte: elaborado pelo autor

Diante das buscas realizadas, observa-se que os sufixos -eira e -eiro formam diferentes palavras, esses sufixos, derivados do latim -arius que carregam diversos significados (VIARO, 2008), sua ocorrência no português serve para a indicar profissão, recipiente, objeto adequado para algo, planta, naturalidade ou qualidade (PRIBERAM, 2008). Entretanto as ocorrências que permanecem no processo de tradução, como observado na figura 5 e tabela 4, correspondem a nomes de plantas da caatinga, evidenciando que o contexto de ocorrência da palavra tende a definir a tradução do seu sentido, desse modo a estruturação do sentido de elementos vegetais

como a Catingueira, ganha uma complementação de sentido através da descrição intratextual ao converter o termo para ‘Catingueira tree’.

Os resultados obtidos das buscas realizadas nessa seção denotam possibilidades de exploração considerando uma de diferentes possibilidades em um *corpus* bilíngue. Observando as escolhas realizadas para traduzir elementos da vegetação, é possível determinar traços de estrangeirização (VENUTI, 2021) no *texto* de chegada, em relação ao *texto* de partida, bem como localizar de maneira eficiente ocorrências de tradução que apagam o estrangeiro e as características particulares a um determinado espaço ou ambiente, como a vegetação da Caatinga observada em *Vidas Secas/Barren Lives*.

### Considerações Finais

Diante do exposto, é possível observar que a ferramenta *LancsBox*®, ao condensar as diferentes funções em um único aplicativo, permite trabalhar com *corpora* com segurança, realizando buscas por diferentes estratégias e obtendo resultados quantitativos que, em conjunção com a análise qualitativa, possibilitam a obtenção de resultados satisfatórios na pesquisa linguística. Nesse sentido, a discussão de estratégias de exploração realizadas em uma pesquisa de área específica da Linguística utilizando uma ferramenta como o *LancsBox*® busca contribuir com a elaboração de estratégias por pesquisadores que, também, utilizam a ferramenta explorada nesse artigo, bem como ferramentas similares de uso mais recorrente.

Observando a utilização em um *corpus* bilíngue, o *LancsBox* mostra-se uma opção viável, considerando a possibilidade de realizar buscas por termos ou expressões regulares. As buscas podem ser realizadas em uma ou em ambas as línguas dos textos do *corpus*, possibilitando obter resultados dinâmicos para análise linguística considerando os objetivos das pesquisas em desenvolvimento. Desse modo, a ferramenta, com o devido manuseio, pode ser um suporte importante para diferentes pesquisas realizadas na Linguística Aplicada.

Diante da possibilidade de buscas em um *corpus* bilíngue, as buscas realizadas no *LancsBox* mostram-se como uma opção viável para realizar comparações e análises no campo dos Estudos da Tradução, permitindo utilizar a ferramenta para buscar terminologias, sufixos, prefixos, radicais e termos que ocorrem em uma ou ambas as línguas, bem como frases e construções em *corpora* para observações distintas. Portanto, nos experimentos apresentados é possível observar desde as estratégias usadas para lidar com itens culturais específicos que causam opacidade no *texto*

de chegada, como as consequências das estratégias adotadas pensando nas redes de significados relacionadas à língua e à cultura do *texto* de partida. Tais resultados podem ser utilizados para refletir sobre estratégias de tradução, projeto de tradução e tendências de tradução observáveis como domesticação e estrangeirização e o que isso representa para o produto da tradução.

#### CRedit

Reconhecimentos: Não é aplicável.

Financiamento: Não é aplicável.

Conflitos de interesse: Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.

Aprovação ética: Não é aplicável.

Contribuições dos autores:

Conceitualização, Curadoria de dados, Análise formal, Aquisição de financiamento, Investigação, Metodologia, Administração do projeto, Recursos, Software, Supervisão, Validação, Visualização, Escrita - rascunho original, Escrita - revisão e edição: MARCELINO, João Gabriel Carvalho.

#### Referências

"-eiro", in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2021, <https://dicionario.priberam.org/-eiro> [consultado em 03-02-2023].

ALUÍSIO, Sandra Maria; ALMEIDA, Gladis Maria de Barcellos. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. *Calidoscópio*, [S. l.], v. 4, n. 3, p. 156–178, 2021. Disponível em: <<http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>>. Acesso em: 1 nov. 2021.

BERBER SARDINHA, Tony. *Corpora* eletrônicos na pesquisa em tradução. *Cadernos de Tradução*. V. 1, No. 9, 2002, p. 15-59. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/5980>>. Acesso em: 18 de out. de 2021.

BERMAN, Antoine. *A tradução e a letra ou o albergue do longínquo*. Tradução de Marie-Hélène C. Torres, Mauri Furlan, Andreia Guerini. 2. ed. Tubarão: Copiart; Florianópolis: PGET/UFSC, 2013.

BREZINA, V., MCENERY, T. & WATTAM, S. (2015). *Collocations in context*: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.

BREZINA, V., TIMPERLEY, M., & MCENERY, A. (2018). *#LancsBox v. 4.x*. [software package].

BREZINA, V., WEILL-TESSIER, P., & MCENERY, A. (2021). *#LancsBox v. 6.x*. [software package]

BRITTO, Paulo Henriques. *A tradução literária*. 2. Ed. Rio de Janeiro: Civilização Brasileira, 2016.

CÂNDIDO, Antônio. *Ficção e Confissão: ensaios sobre Graciliano Ramos*. 4. Ed. Rio de Janeiro: Ouro sobre azul, 2012.





CUNHA, E. L. T. P.. *Contributions to the computational processing of diachronic Linguistic Corpora*. Tese (Doutorado em Linguística / Ciência da Computação) - Universiteit Leiden, Holanda, p. 221. 2020.

FRANCO-AIXELÁ, Javier. Itens Culturais-Específicos em Tradução. Tradução de Mayara Matsu Marinho e Roseni Silva. *In-Traduções*, Florianópolis, v. 5, n. 8, p. 185-218, Jan/jun., 2013. Disponível em: <http://incubadora.periodicos.ufsc.br/index.php/intraducoes/article/viewFile/2119/2996>

JURAFSKY, Daniel; MARTIN, James H. *Speech and Language Processing: An Introduction to Natural Language: Processing, Computational Linguistics, and Speech Recognition*. 3 ed. New Jersey: Prentice Hall, 2022.

KNIGHT, Dawn. The future of multimodal corpora. *Rev. bras. linguist. apl. [online]*. 2011, vol.11, n.2, pp.391-415. ISSN 1984-6398. Disponível em: <http://dx.doi.org/10.1590/S1984-63982011000200006>

McENERY, Tony; HARDIE, Andrew. *Corpus Linguistics*. Cambridge: Cambridge University press, 2012.

OLOHAN, Maeve. *Introducing corpora in translation studies*. New York: Routledge, 2004.

OUSTINOFF, Michaël. *Tradução: história, teorias e métodos*. Tradução de Marcos Marcionilo. São Paulo: Parábola, 2011.

RAMOS, Graciliano. *Barren Lives*. Tradução de Ralph Edward Dimmick. USA: University of Texas Press, 1999.

RAMOS, Graciliano. *Vidas Secas*. 139 ed. Rio de Janeiro: Record, 2018.

VENUTI, Lawrence. *A invisibilidade do Tradutor: uma história da tradução*. Tradução de Laureano Pellegrin... [et al.]. São Paulo: Editora Unesp, 2021.

VENUTI, Lawrence. *Escândalos da Tradução: por uma ética da Diferença*. Tradução de Laureano Pellegrin, Lucinéia Marcelino Villela, Marileide Dias Esqueda, Valéria Biondo. São Paulo: Editora Unesp, 2019.

VIARO, Mário Eduardo. *A formação do significado agentivo de -eiro*. In: XVI Congreso internacional de la ALFAL, 2011, Alcalá de Henares. *Actas del XVI Congreso Internacional de La Asociación de Lingüística y Filología*. Alcalá de Henares: Universidad de Alcalá, 2011. p. 2671-2679. Disponível em <http://www.usp.br/gmhp/publ/ViaA5.pdf>

VIEIRA, Renata; STRUBE DE LIMA, V. L. . *Lingüística Computacional: princípios e aplicações*. In: Ana Teresa Martins; Díbio leandro Borges. (Org.). *SBC - Jornadas de Atualização em Inteligência Artificial (JAIA)*. Fortaleza - CE, 2001, v. 3, p. 47-86.

WILLIAMS, Jenny; CHESTERMAN, Andrew. *The Map: a beginner's guide to doing research in translation studies*. Manchester, UK & Kinderhook: St. Jerome Publishing, 2010.