



O uso de ferramentas computacionais na elaboração do *Vocabulário de Eulálio Motta*: *AntConc e FLE_x*

Computational tools used in the elaboration of the vocabulary of Eulálio Motta:
AntConc e FLE_x

Liliane Lemos Santana Barreiros*
Universidade Estadual de Feira de Santana
Feira de Santana, Bahia, Brasil

Resumo: Com o crescente desenvolvimento das tecnologias computacionais, várias ferramentas estão disponíveis com o intuito de facilitar a construção de obras lexicográficas (dicionários, vocabulários, glossários etc.). Neste trabalho, objetiva-se descrever a funcionalidade e as etapas de aplicação de dois programas utilizados para a construção do *Vocabulário* do escritor baiano Eulálio de Miranda Motta: o *AntConc*, para a análise do *corpus*, e o *Fieldworks Language Explorer (FLE_x)*, para a organização do banco de dados e a edição dos verbetes. Adotou-se como base os princípios teórico-metodológicos da Lexicografia moderna (BIDERMAN, 1969; 1974; 1978; 1984; 2001; HAENSCH, 1982; BARBOSA, 2001, entre outros) e da Linguística de Corpus (BEBER SARDINHA, 2004; OTHERO; MENUZZI, 2005). Delimitou-se como *corpus* desta pesquisa 215 textos em prosa, escritos em vida ou publicados postumamente, sendo: 36 textos publicados na coluna *Rabiscos* do jornal *Mundo Novo* (1931 a 1932, Mundo Novo-BA); 17 textos publicados no jornal *O Lídador* (1933 a 1935, Jacobina-BA); 45 textos publicados no jornal *O Serrinhense* (1950 a 1951, Serrinha-BA); 24 textos publicados no jornal *Gazeta do Povo* (1960-1961, Feira de Santana-BA); 43 panfletos escritos de 1949 a 1983 (BARREIROS, P., 2015) e 50 *causos* que compõem *Bahia Humorística*, escritos de 1933 a 1934 (BARREIROS, L., 2016). Entende-se que essa seleção de 215 textos atende aos diversos suportes utilizados por Eulálio Motta para veiculação e divulgação de seus escritos, possibilitando inventariar o vocabulário usado pelo escritor durante um período de mais de 50 anos (de 1931 a 1983). O *Vocabulário* produzido é composto por 700 entradas, organizadas de A a Z, sendo 505 lexias simples, 61 lexias compostas e 134 lexias complexas (POTTIER, 1977), e tem como finalidade contribuir para a preservação de costumes e valores culturais do homem sertanejo, expresso no uso da língua.

Palavras-chave: Eulálio Motta. Vocabulário. *AntConc*. *FLE_x*.

Abstract: With the increasing development of computational technologies, several tools are available to facilitate the construction of lexicographic works (dictionaries, vocabularies, glossaries, etc.). This paper aims to describe the functionality and stages of application of two softwares used to construct the *Vocabulary* of the bahian writer Eulálio de Miranda Motta: *AntConc* for the analysis of the corpus and the *Fieldworks Language Explorer (FLE_x)* for the organization of the database and the editing of the entries. The theoretical-methodological principles of modern Lexicography (BIDERMAN, 1969; 1974; 1978; 1984; 2001; HAENSCH, 1982; BARBOSA, 2001, among others) and a Linguistics Corpus (BEBER SARDINHA, 2004; OTHERO; MENUZZI, 2005). Delimited as the corpus of this research were 215 texts in prose, written in life or published

* Doutora em Língua e Cultura pela Universidade Federal da Bahia. Professora do Departamento de Letras e do PPGEL/MEL da Universidade Estadual de Feira de Santana. lilianebarreiros@uefs.br.

posthumously, as follows: 36 texts published in the column *Rabiscos* of the newspaper *Mundo Novo* (1931 to 1932, Mundo Novo-BA); 17 texts published in the newspaper *O Lídador* (1933 to 1935, Jacobina-BA); 45 texts published in the newspaper *O Serrinhense* (1950 to 1951, Serrinha-BA); 24 texts published in the newspaper *Gazeta do Povo* (1960-1961, Feira de Santana-BA); 43 pamphlets written from 1949 to 1983 (BARREIROS, P., 2015) and 50 *causos* or anecdotes that composed the book *Bahia Humorística*, written from 1933 to 1934 (BARREIROS, L., 2016). It is understood that this selection of 215 texts attends to the various supports used by Eulálio Motta to promote and disseminate his writings, making it possible to inventory the vocabulary used by the writer during a period of more than 50 years (1931 to 1983). The vocabulary produced is composed of 700 entries, organized from A to Z, being 505 simple lexias, 61 compound lexias and 134 complex lexias (POTTIER, 1977), and aims to contribute to the preservation of cultural customs and values of the country man, as expressed in the use of the language.

Keywords: Eulálio Motta. Vocabulary. *AntConc*. *FLEX*.

1 INTRODUÇÃO

O uso dos computadores e das técnicas quantitativas para análise da língua portuguesa iniciou na década de 1960. Neste período, grupos de pesquisadores do Brasil, Estados Unidos e França trabalharam simultaneamente, conforme seus interesses, em três linhas: “1) fins literários e/ou estilísticos; 2) fins linguísticos e 3) informáticos” (BIDERMANN, 1978, p. 64). Até então, as pesquisas em torno da língua, geralmente, eram voltadas para o ensino, e o *corpus* era coletado, registrado em fichas e analisado manualmente. Uma tarefa muito difícil, lenta e de alto custo.

Neste contexto, destaca-se um trabalho pioneiro para a lexicografia brasileira, a tese de doutoramento de Maria Tereza Camargo Biderman, intitulada *Análise Computacional de Fernando Pessoa (Ensaio de Estatística Léxica)*, defendida em 1969, na Universidade de São Paulo – USP. Nos primórdios das discussões, Biderman utilizou o computador na tabulação de dados léxicos, com vistas a uma análise estilística da obra poética de Fernando Pessoa. Dentro do mesmo domínio de conhecimento, em 1974, na USP, ela defendeu a tese de livre-docência *A Categoria do Gênero*, na qual manipulou dados léxicos dos dicionários de frequência das línguas românicas, confrontando os tipos de gênero que existem nas cinco línguas (português, espanhol, francês, italiano e romeno).

A partir de então, com a popularização dos computadores pessoais, no início da década de 1980, aliado ao desenvolvimento da Linguística de Corpus, as pesquisas passaram a priorizar a descrição da língua e a análise de grande quantidade de dados tornou-se mais confiável. Além disso, possibilitou testar as hipóteses, confrontando a teoria com os dados empíricos da língua. De acordo com Berber Sardinha (2004):

A Linguística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (BERBER SARDINHA, 2004, p. 3).

A Linguística de Corpus está atrelada à Linguística Computacional que, por sua vez, é uma área da ciência linguística voltada para o tratamento computacional da linguagem e das línguas naturais (OTHERO; MENUZZI, 2005, p. 25). Este campo do saber estabelece uma interseção entre os conhecimentos desenvolvidos pela engenharia da computação e as diferentes pesquisas tradicionalmente conhecidas em Linguística teórica e aplicada. Como resultado dessa parceria entre os linguistas e os engenheiros da computação, destaca-se a criação de *softwares* para o estudo de determinados fenômenos linguísticos e a sua ocorrência, a partir da constituição de grandes bancos de dados eletrônicos.

Para Biderman (1984), “o advento do computador constituiu uma verdadeira revolução dentro da ciência da informática e da lexicografia em particular” (BIDERMAN, 1984, p. 17). Na segunda edição do livro *Teoria linguística*, ela destaca que no domínio filológico e literário as técnicas computacionais também foram benéficas (BIDERMAN, 2001, p. 87). Contudo, “a Lexicografia e a Terminologia são indubitavelmente as áreas do conhecimento linguístico que mais se beneficiaram com as informações propiciadas pelos *corpora* de língua escrita e falada e pelas ferramentas computacionais” (BIDERMAN, 2001, p. 92).

Os benefícios alcançados certamente são inúmeros. Além da velocidade na execução das atividades e da ampla capacidade de armazenamento de dados, as novas tecnologias permitem ao lexicógrafo coletar, selecionar, registrar, analisar, aperfeiçoar, recuperar os dados e gerar documentos publicáveis com baixo custo. Ressalta-se que essas vantagens não se limitam apenas na execução das etapas de elaboração de grandes dicionários, mas favorecem ao progresso das pesquisas linguísticas de diversas áreas, atrelando produtividade, com qualidade e acessibilidade.

Para a elaboração do *Vocabulário de Eulálio Motta* (BARREIROS, L., 2017), o primeiro passo foi delimitar o *corpus*. A respeito do conceito de *corpus*, Berber Sardinha (2004) analisa algumas definições e destaca a elaborada por Sanchez-Cantos, em 1996, segundo a qual *corpus* é:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (BERBER SARDINHA, 2004, p. 18).

Berber Sardinha (2004, p. 19-20) assinala ainda quatro pré-requisitos para a formação de um *corpus* eletrônico: 1) a origem – o *corpus* deve ser composto de textos autênticos, em linguagem natural; 2) a autenticidade – escritos por falantes nativos; 3) o conteúdo – deve obedecer a critérios estabelecidos pelo criador de modo que o *corpus* coletado corresponda às características desejadas; e 4) a representatividade – deve ter uma extensão representativa.

Neste sentido, delimitou-se como *corpus* para a construção do *Vocabulário de Eulálio Motta* as publicações de textos em prosa, escritos em vida ou publicados postumamente, sendo: 36 textos publicados na coluna *Rabiscos* do jornal *Mundo Novo* (1931 a 1932); 17 textos publicados no jornal *O Lídador* (1933 a 1935); 45 textos publicados no jornal *O Serrinhense* (1950 a 1951); 24 textos publicados no jornal *Gazeta do Povo* (1960-1961); 43 panfletos escritos de 1949 a 1983 (BARREIROS, P., 2015) e 50 *causos* que compõem *Bahia Humorística*, escritos de 1933 a 1934 (BARREIROS, L., 2016). Entende-se que essa seleção de 215 textos atende aos diversos suportes utilizados por Eulálio Motta para veiculação e divulgação de seus escritos, possibilitando inventariar o vocabulário usado pelo escritor durante um período de mais de 50 anos (de 1931 a 1983). Trata-se de textos autênticos, selecionados criteriosamente, que perfazem um total de 85.845 *tokens* (número total de palavras do *corpus*) e 12.861 *types* (número de palavras do *corpus* contadas apenas uma vez), cumprindo assim o requisito de representatividade.

Após a etapa de seleção dos textos, passou-se à etapa de edição e, em seguida, à compilação propriamente dita. Os textos foram salvos em arquivos separados em formato Word, PDF e TXT e nomeados conforme o código estabelecido por L. Barreiros (2017). Esta nomeação seguiu um padrão para facilitar a recuperação posterior de cada texto.

Atualmente, com o crescente desenvolvimento das tecnologias computacionais, várias ferramentas estão disponíveis com o intuito de facilitar a construção de trabalhos lexicográficos. Após análise e testes com alguns *softwares* que poderiam auxiliar na construção do *Vocabulário de Eulálio Motta*, foram escolhidos dois programas: o *AntConc*, para a análise do *corpus*, e o *Fieldworks Language Explorer (FLEX)*, para a edição e a organização dos verbetes, pois demonstraram eficiência, principalmente, em virtude da flexibilidade, visto que se adaptam às necessidades específicas de cada produção. A seguir, descrevem-se a funcionalidade e as etapas de aplicação de cada um dos programas adotados.

2 ANTCONC: SOFTWARE PARA ANÁLISE DE CORPUS

O *AntConc* é um *software* gratuito para análise de *corpus*, que foi desenvolvido pelo pesquisador Laurence Anthony da Faculdade de Ciências e Engenharia da Universidade de Waseda no Japão. Trata-se de uma ferramenta multiplataforma, executável em versões para o *Windows*, *Linux* e *Macintosh*. Seu arquivo tem aproximadamente 4Mb, por isso é considerado leve e dispensa a necessidade de instalação, o que torna a utilização mais fácil mesmo para usuários iniciantes. Optou-se pelo modelo *AntConc 3.4.4*¹, a última versão para o *Windows*, que dispõe em sua *homepage* (cf. Figura 1) de um canal de suporte com manual, tutoriais em vídeo e grupos de discussão traduzidos do japonês para o inglês.

¹ Cf. Anthony (2014). Download disponível em: <http://www.laurenceanthony.net/software/antconc/>. Acesso em: 6 abr. 2017.

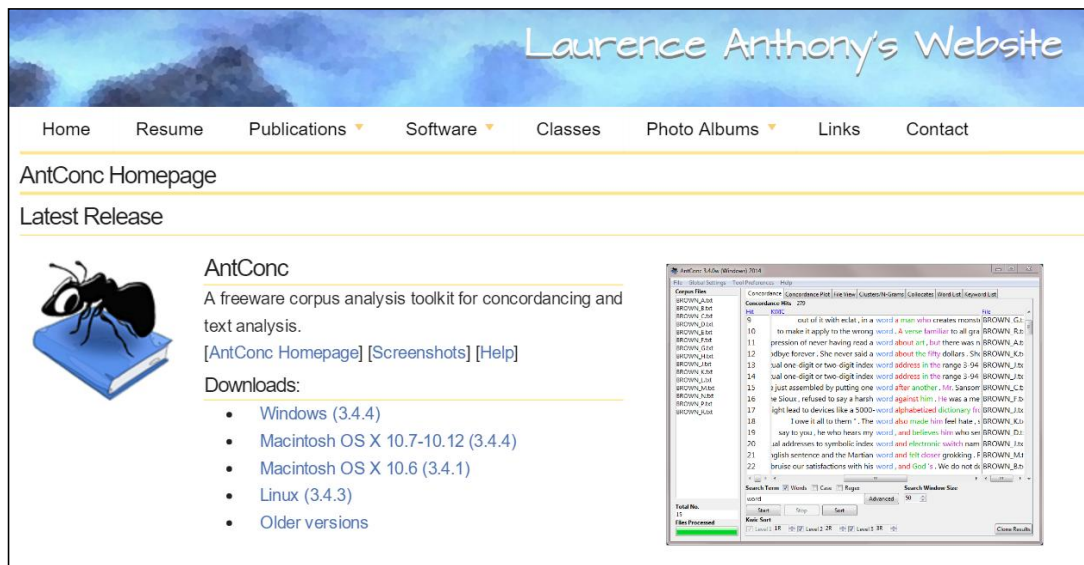


Figura 1 – Homepage do AntConc
Fonte: Anthony (2014).

A interface do *AntConc* é bem simples e, na mesma janela, é possível navegar por diferentes opções de análise, que permitem descobrir como uma palavra ocorre, o quanto ocorre, em que contextos e quais a acompanha, encontrar padrões e variáveis de uso na escrita e fazer levantamento terminológico. Sua praticidade de uso possibilita a extração de listas de palavras (*Word List*), listas de concordâncias (*Concordance*) e de palavras-chaves (*KeyWord*), além de gerar gráficos com os dados analisados. Estas ferramentas são de grande relevância para o linguista, em especial, para o lexicógrafo, pois fornece o conjunto das combinações e das colocações que a palavra pode ter em um determinado *corpus*. Biderman (1984) destaca que:

[...] as concordâncias de texto são um manancial riquíssimo para documentar e ilustrar usos semânticos e gramaticais e atestar o que está ocorrendo de fato na língua, quando se trata de um trabalho sobre Lexicografia contemporânea (BIDERMAN, 1984, p. 22).

Nesta perspectiva, o uso das ferramentas computacionais permite fazer uma avaliação prévia, de forma rápida e econômica, obtendo grande quantidade de resultados em pouco tempo. O que possibilita ao pesquisador ter evidências para provar uma hipótese sobre determinado aspecto linguístico.

Para iniciar o trabalho com o *AntConc*, foi necessário carregar os arquivos com o *corpus* em TXT pelo menu *File*². Para converter os arquivos compilados, utilizou-se o programa *AntFileConverter*³, que é uma ferramenta *freeware* que converte arquivos de PDF

² Todos os itens comentados ao longo do texto estarão destacados em vermelho nas imagens ilustrativas.

³ Disponível em: <<http://www.laurenceanthony.net/software/antfileconverter/>>. Acesso em: 6 abr. 2017.

e Word (DOCX) em texto simples para uso em ferramentas de *corpus*. O *AntFileConverter* está na mesma plataforma do *AntConc* e é indicado pelo programa. Além de ser rápido e de fácil manuseio, ele salva os arquivos em TXT direto na pasta de origem, otimizando a inserção dos mesmos no programa de análise. Essa tarefa é feita pela opção *Open File(s)*, que permite abrir os arquivos individualmente, ou pela opção *Open Dir*, quando for carregar uma pasta completa. Os nomes e a quantidade dos arquivos selecionados aparecerão no quadro *Corpus Files*, à esquerda da tela, como se pode ver na Figura 2.

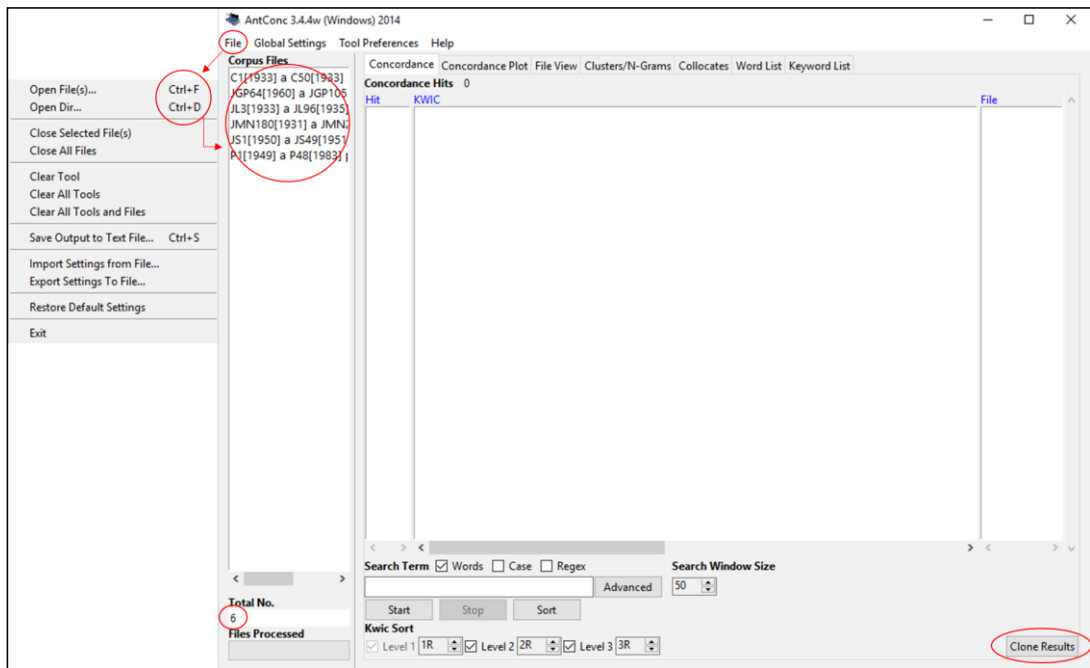


Figura 2 – Tela inicial do *AntConc*

Fonte: Elaborada pela pesquisadora.

Ainda no menu *File*, tem-se a opção de selecionar e excluir arquivos, salvar os resultados das análises realizadas, importar e exportar os dados e restaurar a configuração original.

O botão *Clone Results*, à direita, no final da tela, permite criar uma cópia dos resultados para que possam ser comparados, em janelas independentes. Essa função é importante, quando se tem interesse em confrontar o conjunto dos dados obtidos.

Antes de ativar qualquer função de análise no *AntConc*, é preciso configurar a língua no menu *Global Settings* (cf. Figura 3). Como o *corpus* está escrito em língua portuguesa, a melhor opção é *Western Latin 1*. Ainda no menu das configurações globais (*Global Settings*), é possível editar as opções de cor para realce na análise, a extensão do arquivo a ser aberto (TXT, html, xml ou ALL – todos esses formatos), o tipo e o tamanho da fonte.

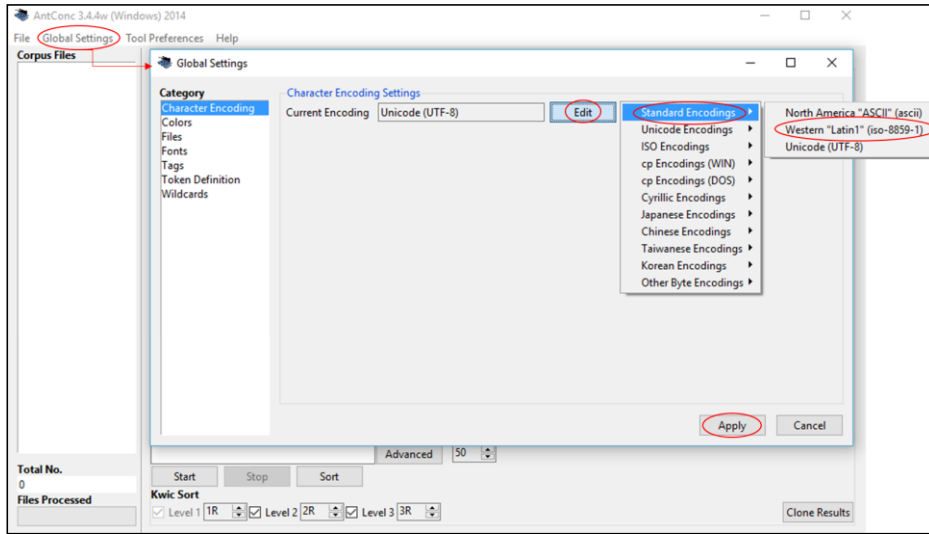


Figura 3 – Configuração do menu *Global Settings* do *AntConc*
Fonte: Elaborada pela pesquisadora.

É importante lembrar também de configurar o menu *Tool Preferences* (cf. Figura 4). Nele, deve-se marcar no campo *Other Options* a caixa *Treat all data as lowercase* (tratar todos os dados como minúsculas) das categorias *Clusters/N-Grams*, *Collocates*, *Word List* e *Keyword List*. Caso contrário, o programa irá diferenciar as letras iniciais entre maiúsculas e minúsculas da mesma palavra, o que causará problemas de exaustividade nos resultados, gerando duas entradas. A depender do objetivo da análise e do tipo de *corpus* trabalhado, essa configuração torna-se desnecessária.

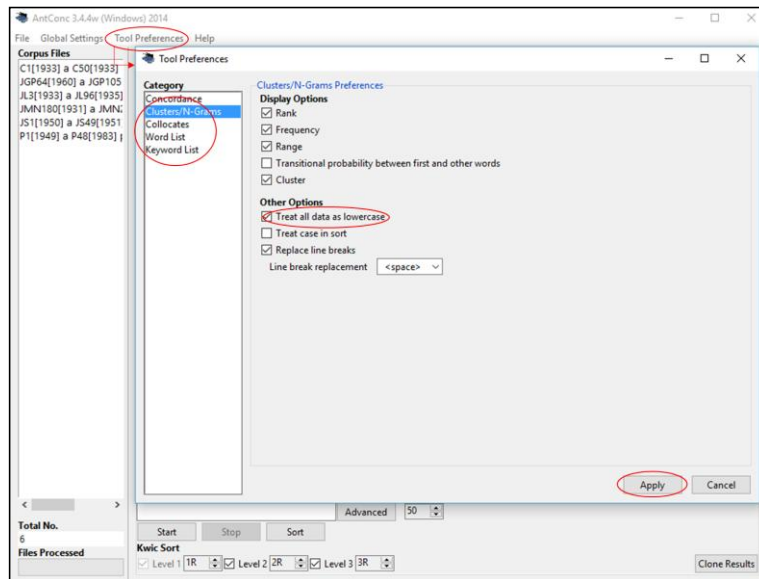


Figura 4 – Configuração do menu *Tool Preferences* do *AntConc*
Fonte: Elaborada pela pesquisadora.

A seguir, descrevem-se as principais ferramentas *Concordance*, *Concordance Plot*, *File View*, *Clusters/N-Grams*, *Collocates*, *Word List* e *Keyword List* e suas respectivas funções dentro do programa *AntConc*.

A primeira ferramenta do menu principal é *Concordance* – ela mostra os resultados da pesquisa em um formato ‘KWIC’ (*KeyWord In Context*), o que permite ver como palavras e frases são comumente usadas em um *corpus*. O resultado é apresentado em três colunas. Na primeira, quantificam-se as ocorrências, na segunda, têm-se as linhas de concordância com destaque para as *palavras-chave* no contexto, e na terceira, a identificação do arquivo em que se encontra. A título de exemplo, buscou-se no *corpus* de Eulálio Motta quantas vezes o escritor usou a forma <Liota> para se referir a si mesmo, como mostra a Figura 5:

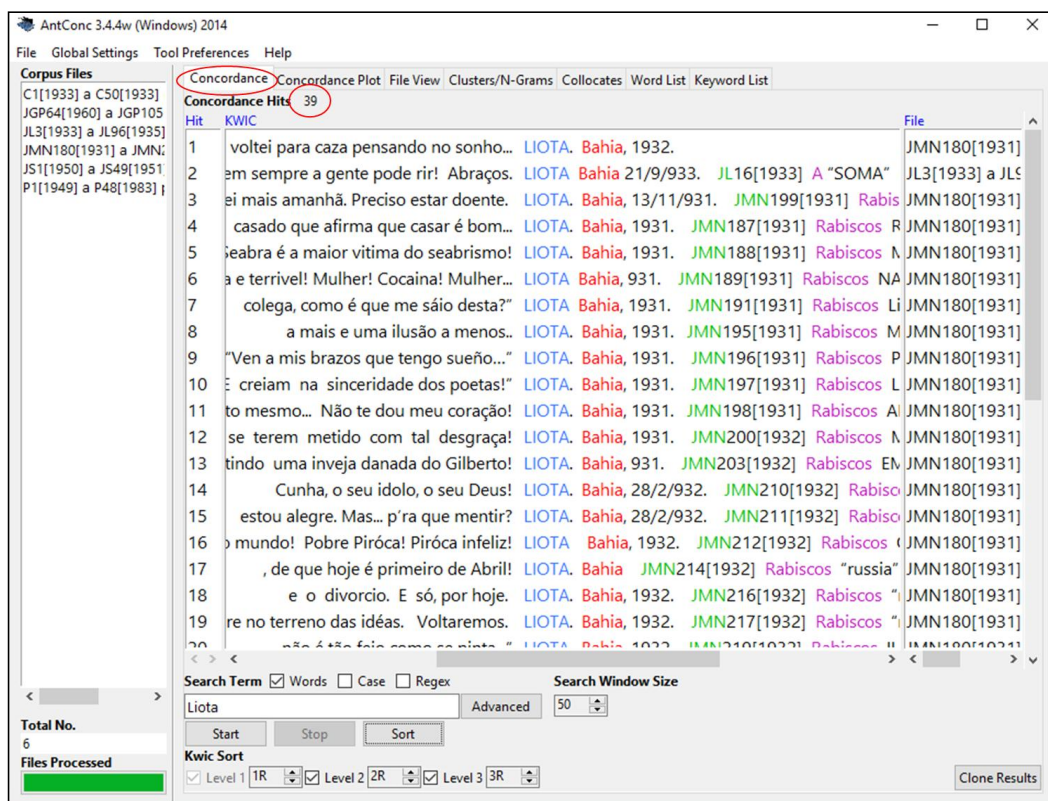


Figura 5 – Resultado da busca por <Liota> na ferramenta *Concordance* do *AntConc*
Fonte: Elaborada pela pesquisadora.

Com a ferramenta *Concordance* foi possível localizar entre os 215 textos que compõem o *corpus* deste trabalho, 39 ocorrências de <Liota>, nos respectivos contextos, com a indicação da data e do código de identificação do arquivo. Com os resultados, foi possível inferir que Eulálio Motta assinou como Liota no período de 1931, 1932 e 1933, apenas nos jornais *Mundo Novo* e *O Lيدador*. A ampliação total do contexto e o acesso ao texto

completo podem ser feitos clicando sobre a palavra-chave desejada ou acionando o recurso *File View*.

Concordance Plot é uma ferramenta de fotografia de concordância, que possibilita visualizar os resultados da pesquisa em forma de gráfico, semelhante a um código de barras, que mostra a distribuição, no arquivo em questão, da palavra que se está analisando. Ela permite ver quais os arquivos incluem a palavra pesquisada e onde aparecem no texto. O número de ocorrências e o comprimento de cada texto é mostrado à direita do gráfico. Este, por sua vez, pode ser ampliado ou reduzido usando o botão *Plot Zoom*. Para exemplificar, buscou-se por <Mundo Novo> e o resultado apresentou ocorrências em todos os arquivos, como se vê na Figura 6:

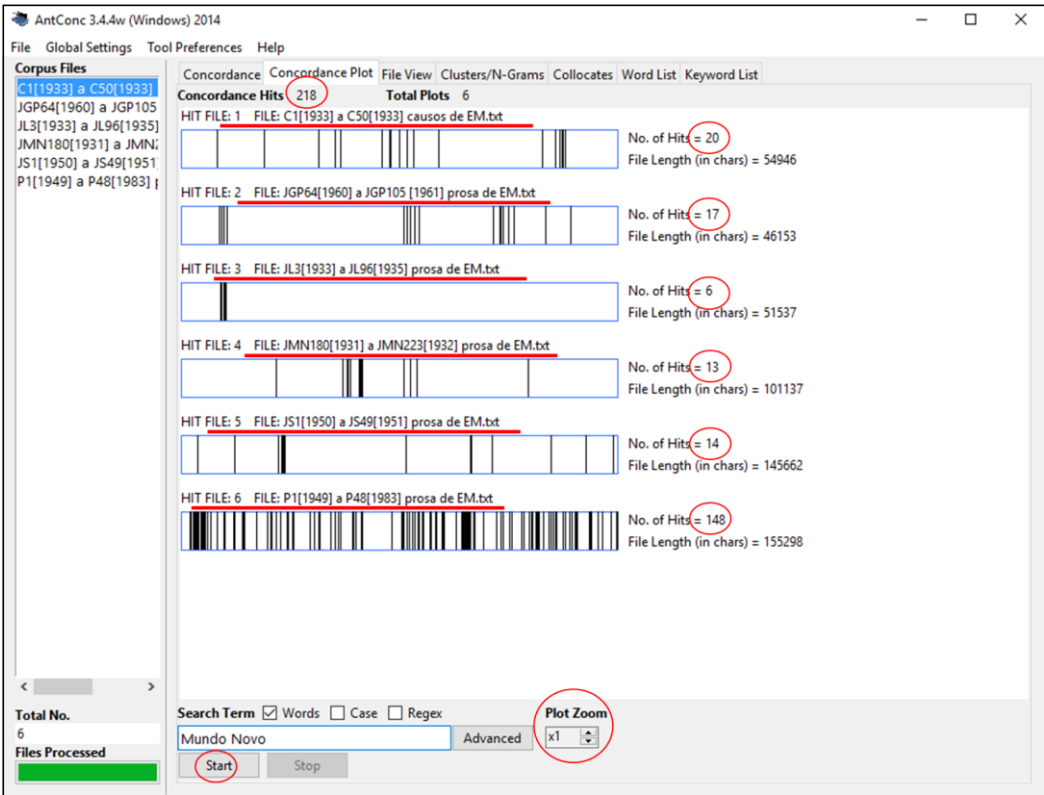


Figura 6 – Resultado da busca por <Mundo Novo> na ferramenta *Concordance Plot* do *AntConc*
Fonte: Elaborada pela pesquisadora.

Com a ferramenta *Concordance Plot*, identificaram-se 218 ocorrências, sendo 6 no jornal *O Lidador*, 13 no jornal *Mundo Novo*, 14 no jornal *O Serrinhense*, 17 no jornal *Gazeta do Povo*, no 20 nos causos e 148 nos panfletos. Ao clicar em cada linha do gráfico, o usuário é remetido ao contexto da ocorrência. Com o auxílio da função *Clusters/N-Grams*, que será explicada mais adiante, identificaram-se 2 ocorrências grafadas com hífen <Mundo-Novo> e uma com acento circunflexo <Mundo Nôvo>.

A função *File View* mostra o texto bruto de arquivos individuais. Isto permite investigar com mais detalhes os resultados gerados pelas ferramentas *Concordance* e *Concordance Plot*, pois as diferentes ocorrências ficam sinalizadas. Neste caso, se uma palavra tiver sido especificada, os resultados serão destacados em todo o texto. Também é permitido alterar a busca. Com o botão *Hit Location* é possível saltar pelas ocorrências sem precisar percorrer todo o arquivo. A Figura 7, exemplifica a busca por <Mundo Novo> no arquivo dos panfletos (P1[1949] a P48[1983]), nos quais registram-se 148 ocorrências:

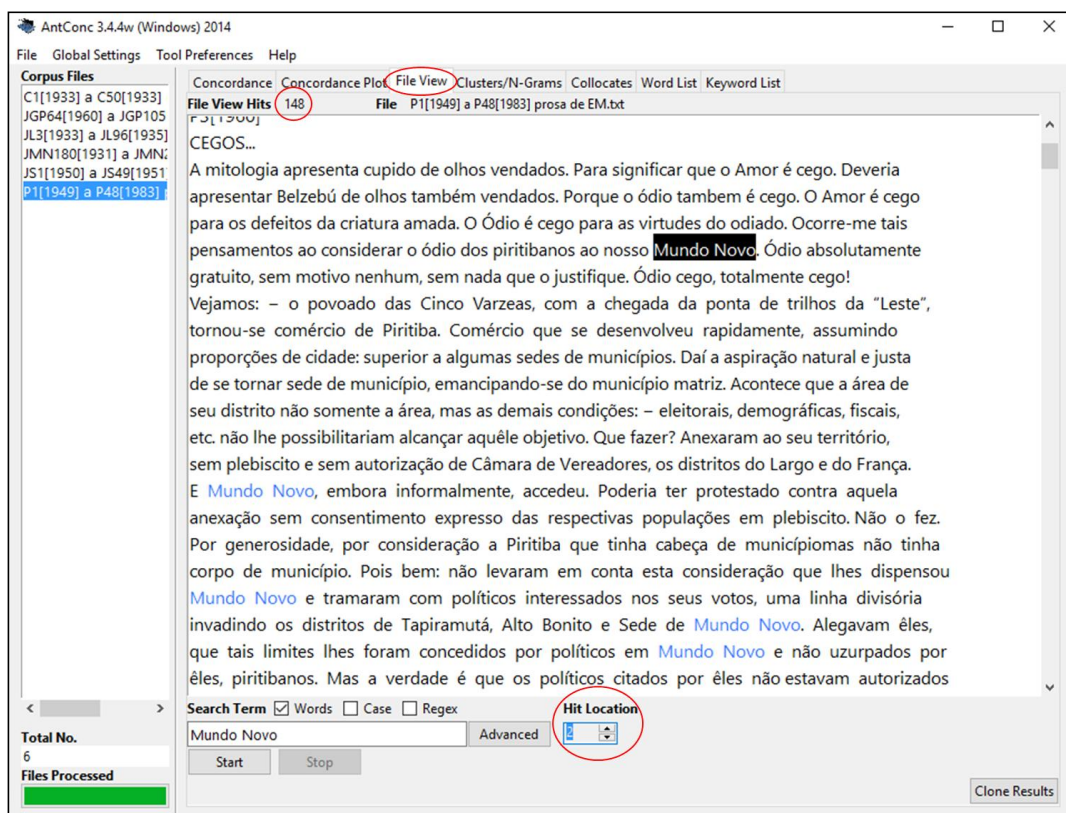


Figura 7 – Resultado da busca por <Mundo Novo> na ferramenta *File View* do *AntConc*
Fonte: Elaborada pela pesquisadora.

A opção *Clusters/N-Grams* do modelo *AntConc* 3.4.4 acopla as funções *Clusters* e *N-Grams* das versões anteriores. Esta ferramenta gera uma lista com o resultado da pesquisa, conforme a ordem escolhida (alfabética, de frequência, de probabilidade ou de terminações). A ordem dessas listas também pode ser invertida, ativando a opção *Invert Order*. Além dessas possibilidades, *Clusters/N-Grams* agrupa os resultados gerados na *Concordance* ou *Concordance Plot* com as palavras mais próximas à esquerda (*On Left* – busca padrão do programa) ou à direita do termo de pesquisa (*On Right*), de acordo com a opção assinalada. Pode-se selecionar o comprimento mínimo e máximo (número de palavras) em cada *cluster* e a frequência mínima de *clusters* exibidos, o que possibilita localizar

expressões comuns em todo o *corpus*. Este procedimento é bem proveitoso quando se trabalha com o levantamento de lexias compostas e complexas.

Na Figura 8, realizou-se uma busca por <integralista>, utilizando a ferramenta *Clusters/N-Grams*, que gerou uma lista por ordem de frequência de 27 ocorrências, com a indicação das 4 palavras mais próximas à direita.

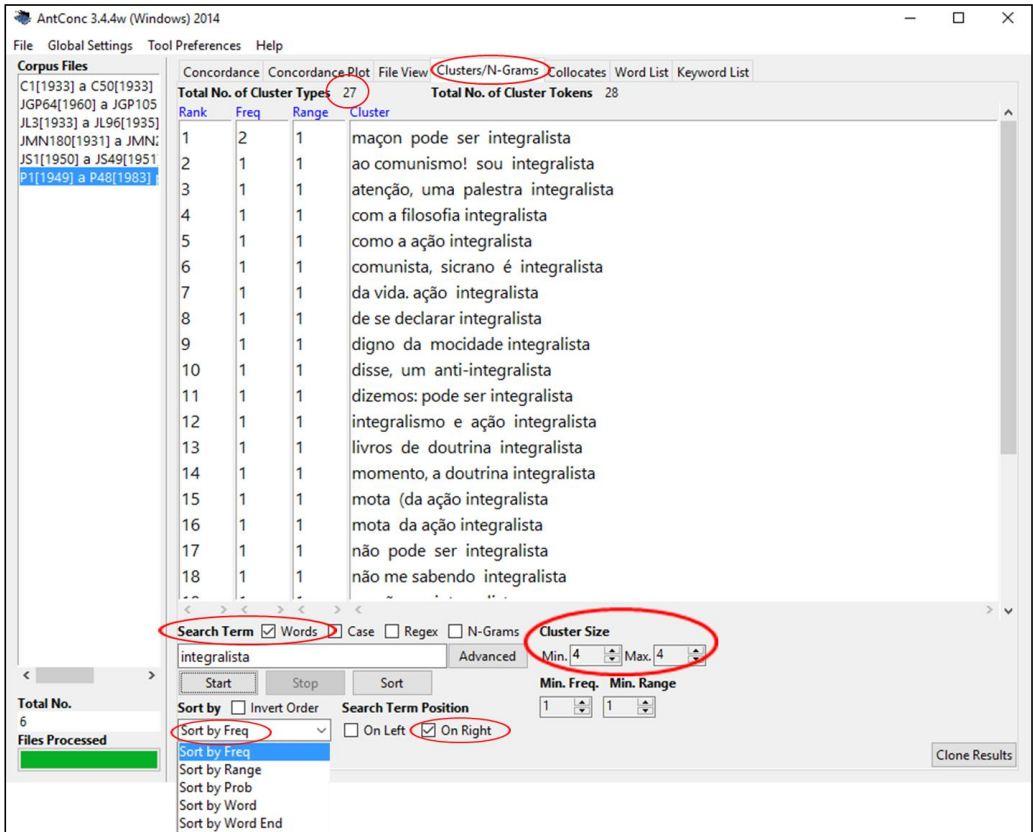


Figura 8 – Resultado da busca por <integralista> na ferramenta *Clusters/N-Grams* do *AntConc*
Fonte: Elaborada pela pesquisadora.

Observa-se que a busca do *Search Term* foi feita por palavras, o que é o padrão. No entanto, é possível fazer por *seqüência de caracteres* desativando a opção *Words*, ou por meio de uma busca *sensível a maiúsculas e minúsculas* ao ativar a opção *Case* ou usando expressões regulares cheias (com caracteres) ativando a opção *Regex*. Estas opções de filtro para a busca estão disponíveis em todas as opções do *AntConc*.

A próxima ferramenta é *Collocates*. Ela permite investigar padrões não sequenciais na língua. Sua função principal é gerar listas ordenadas das palavras próximas ao termo pesquisado, chamadas de colocados. Estes são ordenados pela frequência total, pela frequência à esquerda e à direita do termo de pesquisa e pelo início e final da palavra. Também apresentam uma média estatística, gerada automaticamente pelo programa, que mede o nível de relação entre o termo pesquisado e o colocado. Além disso, pode-se

selecionar a extensão de palavras à esquerda e à direita do termo de pesquisa para localizar os colocados e a sua frequência mínima exibida. Se, por exemplo, for necessário um intervalo de apenas uma palavra para ver as quais aparecem diretamente à direita do termo de pesquisa, é necessário marcar a opção *Same*, para manter o tamanho mínimo e máximo do intervalo. Nesta ferramenta, pesquisou-se pela variação <cuma> para ver as possíveis colocações que integra como base, o que se mostra nas Figuras 9 e 10.

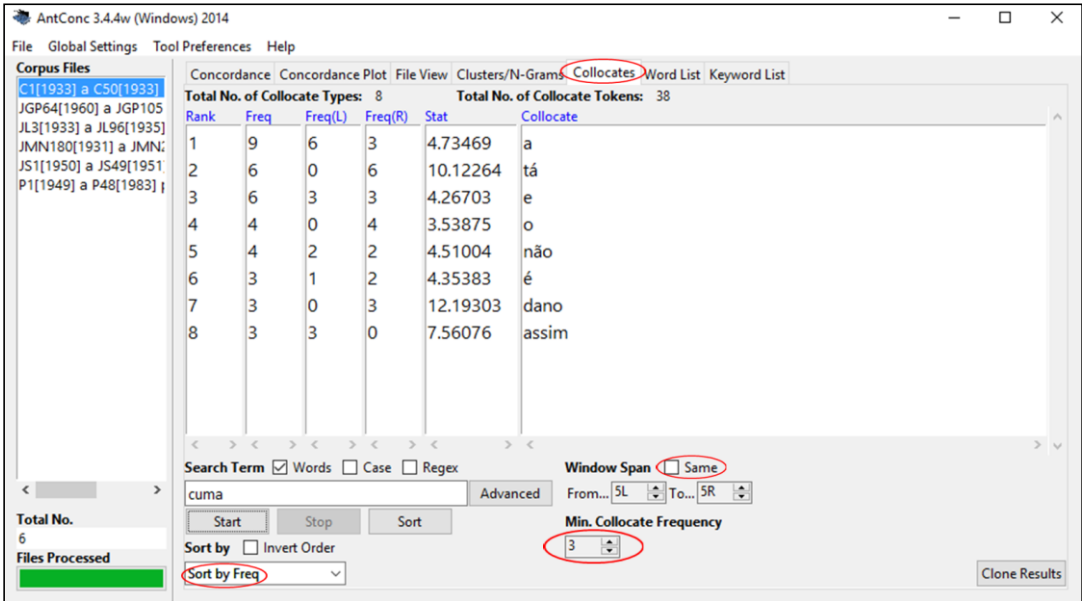


Figura 9 – Resultado da busca por <cuma> na ferramenta *Collocates* do *AntConc*
Fonte: Elaborada pela pesquisadora.

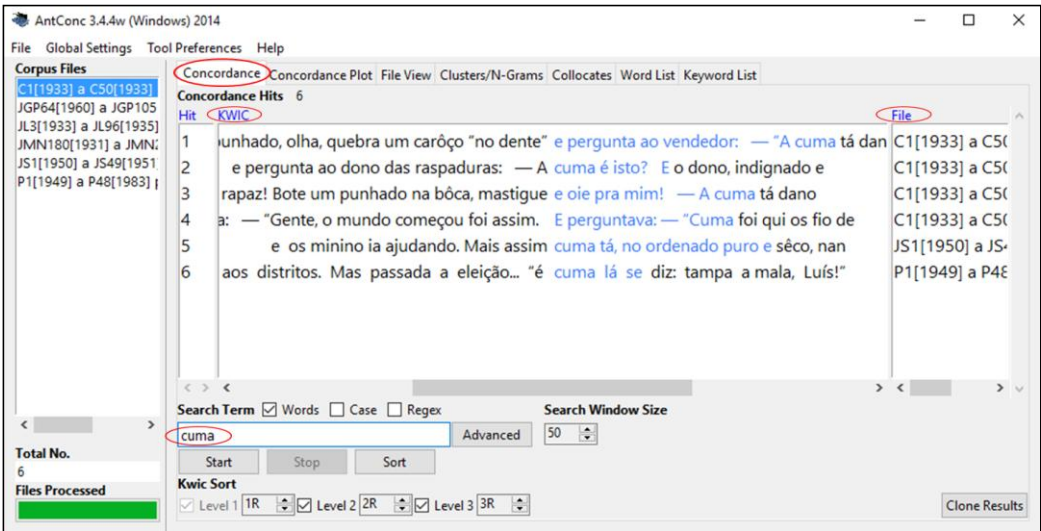


Figura 10 – Resultado da busca por <cuma> na ferramenta *Concordance* do *AntConc*
Fonte: Elaborada pela pesquisadora.

Ao ativar a função *Concordance* visualizou-se o termo buscado nos respectivos contextos como base das colocações: “a cuma tá dano”, “a cuma é isto”, “cuma foi”, “cuma tá” e “cuma lá se diz”. Na coluna *File*, tem-se a identificação da localização no *corpus*. Neste caso, as ocorrências foram encontradas nos causos, no panfleto e no jornal *O Serrinhense*.

Por fim, as ferramentas *Word List* (lista de palavras) e *Keyword List* (listas de palavras-chave). A *Word List*, gera uma lista, ou em ordem alfabética ou por frequência, de todas as palavras que constam nos arquivos selecionados para a análise linguística. Assim, é possível encontrar rapidamente quais palavras são as mais frequentes e quais são as com menor ocorrência em um *corpus*, neste caso, deve-se assinalar a caixa *Invert Order*.

A Figura 11, mostra a lista de palavras gerada pelo *AntConc*, que possui três colunas: 1) *Rank* – indica a classificação das palavras do *corpus* por ordem de frequência; 2) *Freq* – mostra a frequência com que cada palavra ocorre no *corpus*; e 3) *Word* – mostra cada palavra referente aos itens das duas colunas anteriores. A lista é organizada conforme o critério escolhido dentre os três disponíveis: 1) *Sort by Freq* – é uma busca padrão do programa que gera uma lista de palavras por critério de frequência (cf. Figura 12); 2) *Sort by Word* – organiza as palavras por ordem alfabética; e 3) *Sort by Word End* – ordena as palavras segundo suas terminações (sílabas finais). Após selecionar a opção, é só clicar em *Start*.

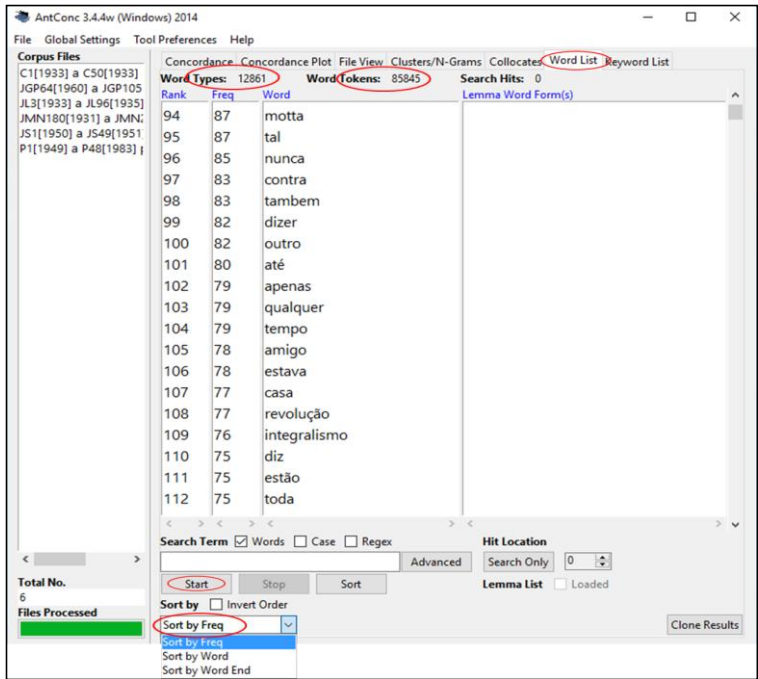


Figura 11 – Resultado da busca por critério de frequência com a ferramenta *Word List* do *AntConc*

Fonte: Elaborada pela pesquisadora.

Como o critério de frequência não foi priorizado para eleger as entradas dos verbetes do *Vocabulário de Eulálio Motta*, adotou-se o critério *Sort by Word* para gerar uma lista de palavras em ordem alfabética. A lista compreende 12.861 tipos de palavras, que abarcam as classes abertas (substantivos, verbos, adjetivos e advérbios) e as classes fechadas (artigos, pronomes, numerais, conjunções etc.), registradas no *corpus* analisado.

A *Keyword List*, por sua vez, gera uma lista de palavras-chave, comparando a frequência das palavras raras do arquivo em análise com a frequência das palavras do *corpus* de referência, o que permite identificar palavras características no *corpus* de estudo como parte de um gênero ou de uma variedade linguística.

Para criar uma lista de palavras-chave, primeiro é preciso definir um *corpus* de referência. Optou-se pelo *Novo Dicionário da Língua Portuguesa*, de Cândido de Figueiredo, por conta do período em que foi publicado (1913) e por estar disponível *online* no formato PDF⁴. Para inserção do *corpus* de referência, destaca-se, na Figura 12, a ordem das ações com a indicação numérica em vermelho.

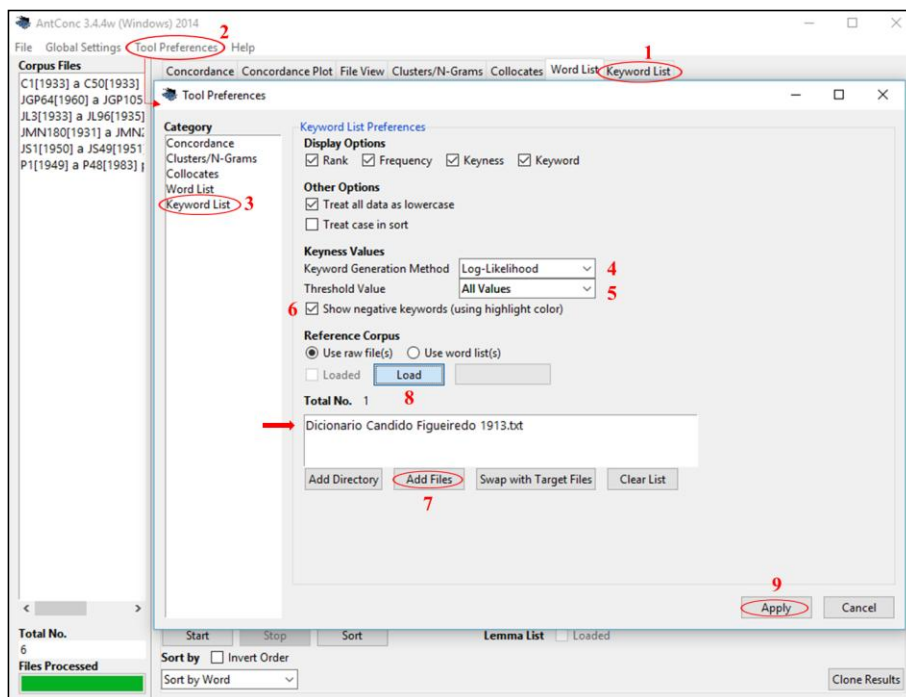


Figura 12 – Inserção do *corpus* de referência na ferramenta *Keyword List* do *AntConc*
Fonte: Elaborada pela pesquisadora.

O primeiro passo é acessar o menu *Keyword List* (1), em seguida, o *Tool Preferences* (2) e a opção *Keyword List* (3). É preciso configurar o método de geração de palavras-chave, ou seja, a medida estatística para calcular a *chavidade* das palavras do arquivo de destino

⁴ Disponível em: <<http://dicionario-aberto.net/dict.pdf>>. Acesso em: 6 abr. 2017.

(4). A configuração padrão recomendada é *Log Likelihood* (log de probabilidade). Depois, escolhe-se um limite para exibição do número de palavras-chave (5) com opções de 100 a 1000, mas o ideal é *All Values* (todos os valores), pois, assim, abarcar todas as ocorrências. É possível escolher se se deseja ou não exibir as *palavras-chave negativas* (6), que são as palavras do arquivo em análise com uma frequência baixa em comparação com a frequência no *corpus* de referência. Neste caso, é interessante assinalar a opção, pois evidencia as marcas de uso no *corpus* de análise. Para adicionar o arquivo com o *corpus* de referência em TXT, seleciona-se a opção *Add Files* (7) e, em seguida, clica-se em *Load* (8) para carregar. Quando se conclui o carregamento do arquivo, clica-se em *Apply* (9) e o programa voltará para a tela inicial em *Keyword List*. Para que o programa processe a lista de palavras-chave do *corpus* em análise, é necessário configurar o campo *Sort by*. Dessa forma, ajusta-se a opção de organização da lista, que pode ser: por frequência, por chave, por ordem alfabética ou por palavra-chave final. Optou-se pela *Sort by Keyword*, que gera a lista de palavras-chave em ordem alfabética para confrontar com a *Word List* também gerada em ordem alfabética. Basta clicar em *Start* e automaticamente a lista é gerada, como demonstrado na Figura 13.

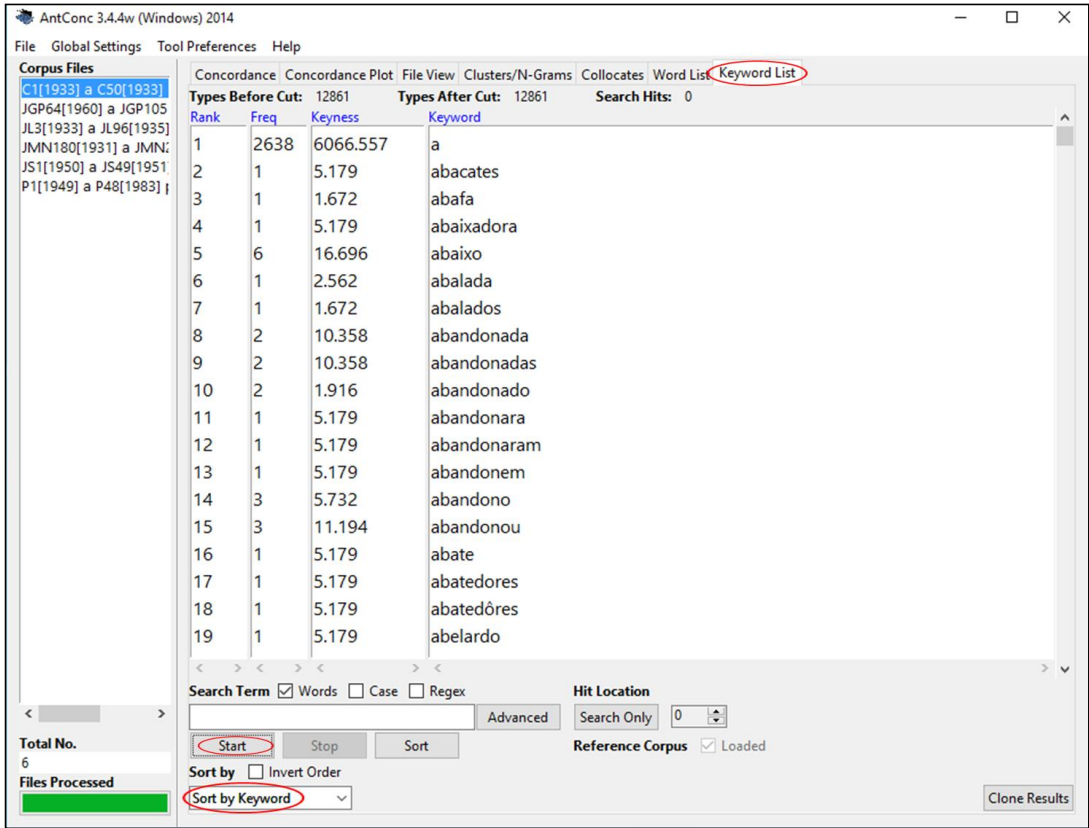


Figura 13 – Resultado da lista de palavras-chave gerada com a ferramenta *Keyword List* do *AntConc*

Fonte: Elaborada pela pesquisadora.

Para salvar os resultados, basta acessar o menu *File* e selecionar a opção *Save Output to Text File*. Para realizar uma nova pesquisa, utilizando o mesmo *corpus*, limpam-se as informações anteriores, com a opção *Clear All Tools and Files* no menu *File*. Vale ressaltar que o programa não diferencia palavras empregadas com mais de um sentido, contabilizando apenas a forma gráfica.

As *palavras-chave negativas*, ou seja, as de baixa frequência, são exibidas na mesma lista, ao final, destacadas em azul. A contagem é reiniciada com elas. Assim, com o *corpus* de Eulálio Motta foi gerada uma lista de 351 palavras de baixa frequência, como na Figura 14.

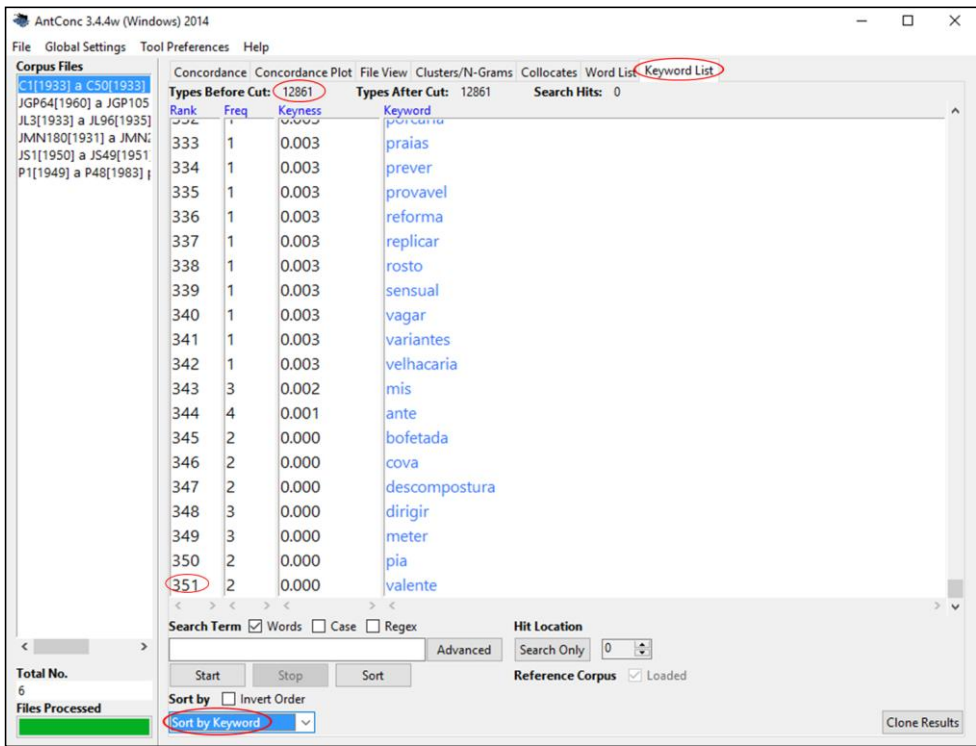


Figura 14 – Resultado da lista de palavras-chave de baixa frequência gerada com a ferramenta *Keyword List* do *AntConc*.
 Fonte: Elaborada pela pesquisadora.

O uso do *AntConc* mostra-se eficiente porque dimensiona o volume de dados do *corpus* em análise, observando-se as ocorrências e as variações de uso da língua. Ele permite explorar o *corpus* em sua totalidade de maneira rápida e eficiente, o que seria muito difícil manualmente. Os resultados obtidos com as listas de palavras são de extrema importância para o próximo passo, que é gerar o *Vocabulário de Eulálio Motta*, pois subsidiarão a escolha das entradas que irão compor a microestrutura do vocabulário.

Para a estruturação do vocabulário, utilizou-se a ferramenta *Fieldworks Language Explorer (FLEx)* que será descrita a seguir.

3 FIELDWORKS LANGUAGE EXPLORER (FLEX) E A ORGANIZAÇÃO DO CORPUS

O *FieldWorks Language Explorer (FLEX)* é um programa gratuito produzido pela SIL – Sociedade Internacional de Linguística. A SIL é uma organização científica sem fins lucrativos, que tem como objetivo principal estudar e documentar línguas minoritárias para traduzir a Bíblia. Para isto, ela desenvolve e disponibiliza diversos *softwares* que auxiliam nas atividades de pesquisa linguística e cultural, alfabetização e educação em língua materna, elaboração de dicionários, tradução comunitária e assistida por computador e produção de mídia vernácula. A SIL foi iniciada em 1934, nos Estados Unidos, como um treinamento para missionários. Atualmente, tem mais de 6.000 membros, em mais de 50 países, e possui *status* consultivo formal da UNESCO e das Nações Unidas⁵.

Entre os *softwares* disponibilizados pela SIL, o *FLEX*⁶ (*homepage* ilustrada na Figura 15) se destaca por ser um banco de dados que foi projetado para auxiliar os linguistas de campo a executar muitas tarefas relacionadas à documentação e à análise da língua em um único programa, como: coletar e registrar informações lexicais, fonológicas, morfossintáticas e antropocultural, interlinearizar textos, criar e publicar dicionários.

Diante das suas possibilidades, o *FLEX* tornou-se útil para a lexicografia, pois ele se adequa às especificidades de cada trabalho, permitindo ao lexicógrafo realizar um banco de dados mais completo com as opções de inserir textos, sons e imagens, exportar os dados em PDF e disponibiliza-los *online*.

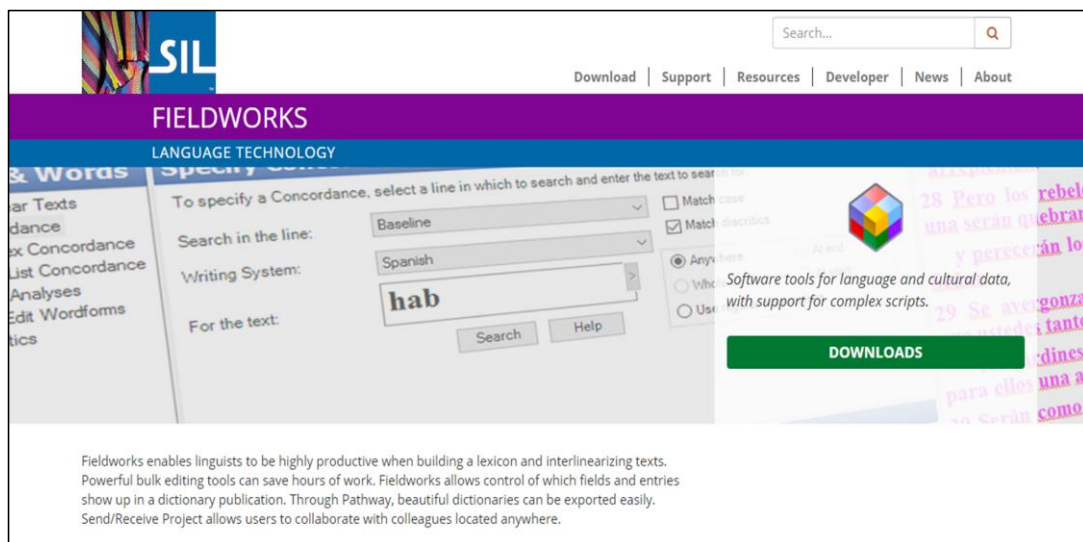


Figura 15 – *Homepage* do *FieldWorks Language Explorer (FLEX)*

Fonte: SIL – *International Society of Linguistics*.

⁵ Dados obtidos na *homepage* da SIL, disponível em: <<https://www.sil.org/about/discover>>. Acesso em: 14 abr. 2017.

⁶ Download disponível em: <<http://software.sil.org/fieldworks/download/>>. Acesso em: 14 abr. 2017.

O *FLEX* é livre para usar, modificar e redistribuir, de acordo com os termos da licença pública geral – *GNU Lesser General Public License*⁷. Ele está sendo constantemente aprimorado. Em 2017, o *FLEX* foi atualizado nos meses de janeiro, março e abril. Esta atualização é feita automaticamente e o usuário não precisa fazer *backup*, desde a versão *FieldWorks 8*. No site *GitHub*⁸, é possível acompanhar como os desenvolvedores desse programa trabalham as suas versões. O *GitHub* é uma plataforma colaborativa de gerenciamento de hospedagem e criação de *softwares*, que permite aos usuários do mundo todo ter acesso, analisar e contribuir com o código das ferramentas. Assim, por meio da página do *FLEX* no *GitHub*, sugerem-se melhorias e correções, visto que a quantidade de contribuições é infinita.

Com o objetivo de analisar a produtividade e a difusão dessa ferramenta, realizou-se uma consulta do *FLEX* no *GitHub*, como se vê na Figura 16, em março de 2017. Verificou-se no menu *Commits* (1), que foram enviadas e aprovadas 2.098 alterações desde a sua criação, sendo que, um mês depois, com a publicação da versão *FieldWorks 8.3.7 RC2*, esse número de participações já estava em 4.384. Essa dinâmica de uso é positiva, pois facilita que falhas sejam encontradas e corrigidas mais rapidamente, além de acréscimos de novas funcionalidades, fazendo com que o programa esteja em constante melhoria e atualização.

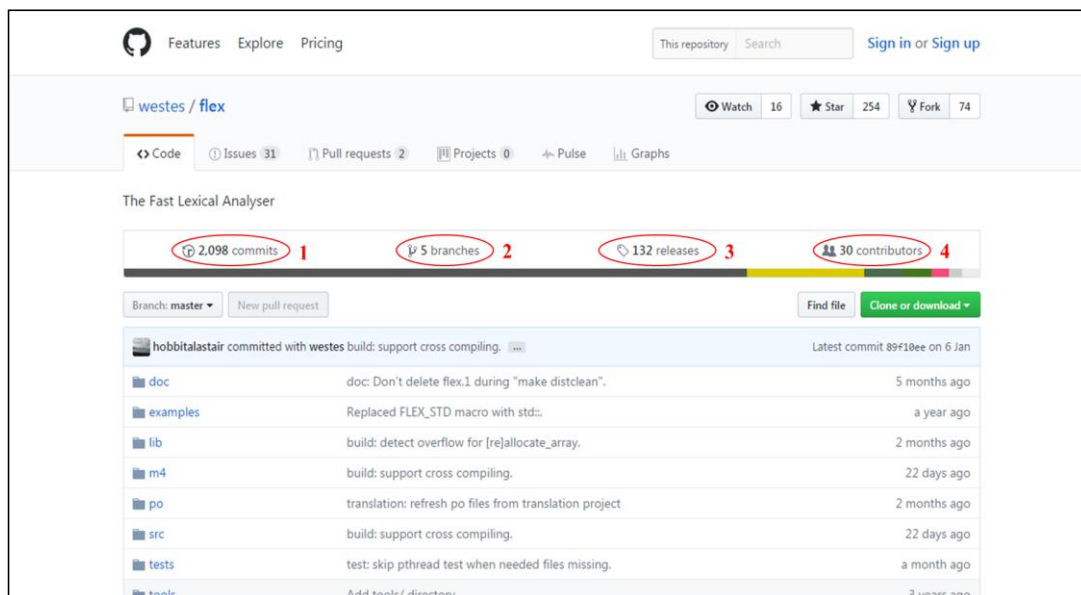


Figura 16 – *Homepage do GitHub/FieldWorks*
Fonte: *GitHub*.

Ainda, na Figura 16, no menu principal do *GitHub*, em *branches* (2), encontram-se disponíveis as cinco árvores utilizadas na divisão do algoritmo da ferramenta. Em *releases*

⁷ Disponível em: <<http://www.gnu.org/licenses/lgpl-2.1.html>>. Acesso em: 23 abr. 2017.

⁸ Disponível em: <<https://github.com/sillsdev/FieldWorks>>. Acesso em: 17 mar. 2017.

(3), estão 132 versões funcionais do sistema que foram liberadas e, até o momento, são 30 colaboradores (4) que cooperam com o código.

O *FLEX* permite que os usuários, com acesso ao mesmo projeto, colaborem com pesquisadores localizados em qualquer lugar, por meio de um processo de sincronização dos dados. Logo, várias pessoas podem participar do mesmo projeto, otimizando o tempo e aumentando a produtividade.

Por todas as possibilidades e vantagens destacadas, adotou-se a versão *FieldWorks 8.3.7 RC2*, atualizada em 6 de abril de 2017, para a realização do *Vocabulário de Eulálio Motta*. Vale ressaltar que este projeto foi cadastrado na plataforma da SIL e está registrado no repositório *Language Depot*⁹, para que seja possível futuramente sincronizar todos os dados obtidos e assim poder ampliar o *corpus* de análise e a equipe de trabalho.

A seguir, descreve-se a estrutura e o funcionamento do programa, a fim de demonstrar a sua aplicação. Após o *download* do *FLEX*, o primeiro passo é a criação de um novo projeto na plataforma e a definição do sistema de escrita, nesse caso, português. Em seguida, o *FLEX* remete o usuário para a tela inicial, como se demonstra na Figura 17.

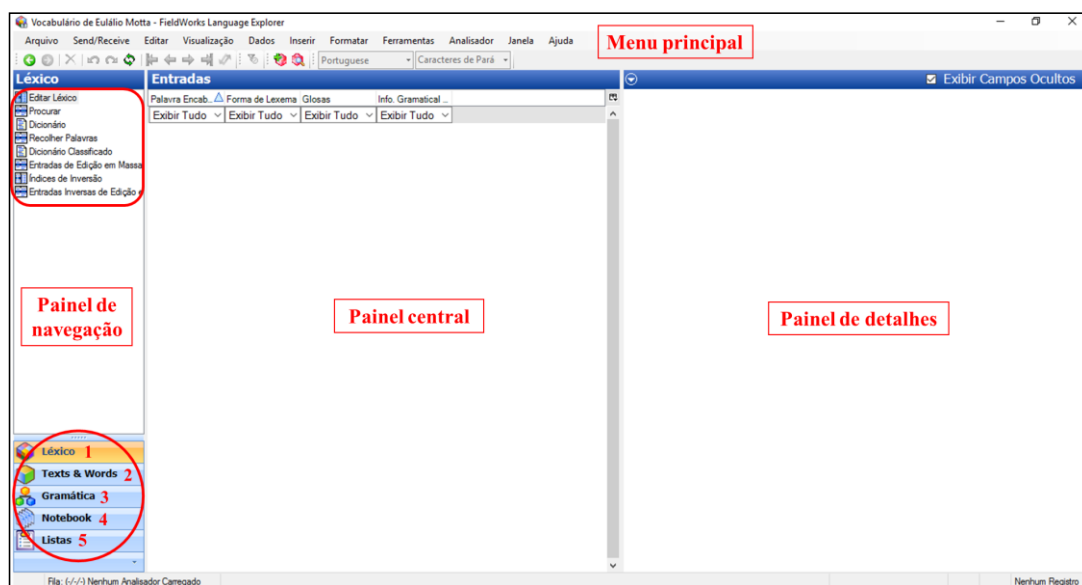


Figura 17 – Tela inicial do *FLEX*

Fonte: Elaborada pela pesquisadora.

Além da barra do menu principal, que navega pelas funções mais gerais do *FLEX*, como abrir ou restaurar um projeto, a janela do programa é dividida em três painéis principais. À esquerda, o painel de navegação, apresenta uma lista, no canto inferior, com as cinco áreas de trabalho – *Léxico*, *Texts & Words*, *Gramática*, *Notebook* e *Listas*¹⁰. Cada

⁹ Disponível em: <<https://public.languagedepot.org/projects/show/br-por-flex?jump=my>>. Acesso em: 28 fev. 2017.

¹⁰ O *FLEX* é originalmente em língua inglesa, por isso, quando se define o sistema de escrita em português, nem todos os campos são traduzidos.

uma tem as suas opções específicas, elencadas no canto superior. No centro, o painel central, e à direita, o painel de detalhes, ambos exibem o campo de edição da ferramenta selecionada à esquerda.

As cinco áreas de trabalho assinaladas na Figura 17 têm as seguintes finalidades: *Léxico* (1) é o espaço onde se constroem os verbetes. Possui campos pré-definidos, que podem ser visualizados ou ocultados de acordo com a configuração adotada. *Texts & Words* (2) é o local onde são armazenados todos os textos transcritos. *Gramática* (3) é o recurso que permite inserir a classificação gramatical das unidades lexicais armazenadas. *Notebook* (4) exhibe todos os dados antropológicos registrados, sendo possível acrescentar informações. Nas versões anteriores do *FLEX*, essa função era desenvolvida por outro programa. Por fim, *Listas* (5) permite a edição de opções de lista para determinados campos no banco de dados, gerada a partir dos dados inseridos durante a configuração do produto.

Para iniciar a construção do *Vocabulário de Eulálio de Motta* na plataforma do *FLEX* foi necessário inserir o *corpus* da pesquisa na área *Texts & Words*, como se mostra na Figura 18, na subseção *Textos Interlinearizados* (1), pelo atalho ou em *Inserir* no menu principal (2).

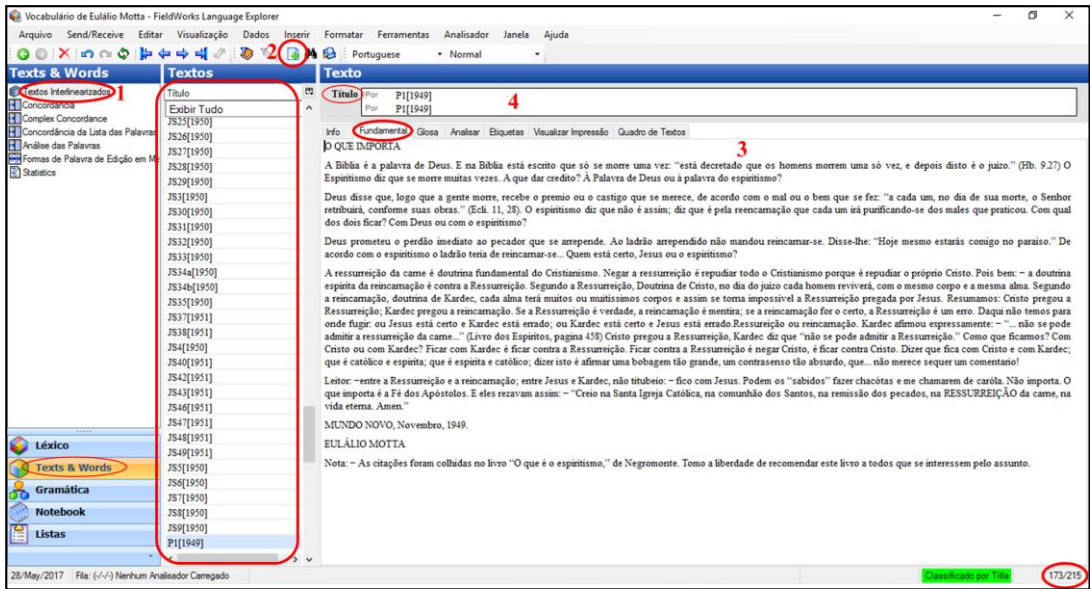


Figura 18 – Inserção do *corpus* da pesquisa na área *Texts & Words* do *FLEX*

Fonte: Elaborada pela pesquisadora.

Uma vez fixados na aba *Fundamental* (3), do terceiro painel, o *corpus* já está salvo. O título pode ser inserido na aba específica (4) ou em *Info*, onde se tem os campos para registrar todos os metadados do texto (fonte, gênero, comentários, participantes, localização e dados da coleta). A data de criação e de modificação dos dados são atualizadas automaticamente. Os metadados dos 215 textos que compõem o *corpus* desta análise foram editados individualmente, como se vê na Figura 19.

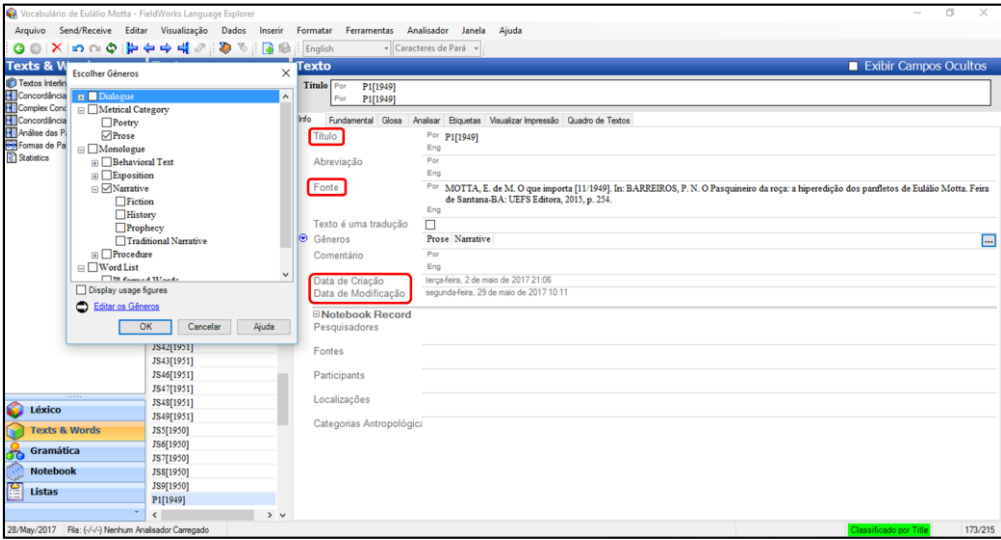


Figura 19 – Edição dos metadados do *corpus* da pesquisa na área *Texts & Words* do FLEX
Fonte: Elaborada pela pesquisadora.

Ainda nessa área de trabalho, *Texts & Words*, é possível editar, extrair a concordância da lista de palavras dos textos conforme sua ocorrência, fazer diversos tipos de buscas e ter acesso à estatística geral. Após a inserção do *corpus*, iniciou-se a criação das entradas lexicais, as quais foram escolhidas a partir de uma pré-seleção da *Word List* gerada pelo *AntConc*. O FLEX possibilita adicionar entradas de quatro maneiras diferentes. Para a organização do *Vocabulário de Eulálio Motta*, elegeram-se duas opções que mais se adequaram à proposta deste trabalho em compor um vocabulário de lexias simples, compostas e complexas. A primeira forma de registro das entradas adotada foi pela área de trabalho *Léxico*, como se demonstra na Figura 20. *Editar léxico* é a principal ferramenta de edição do FLEX, pois permite registrar várias informações para cada entrada lexical.

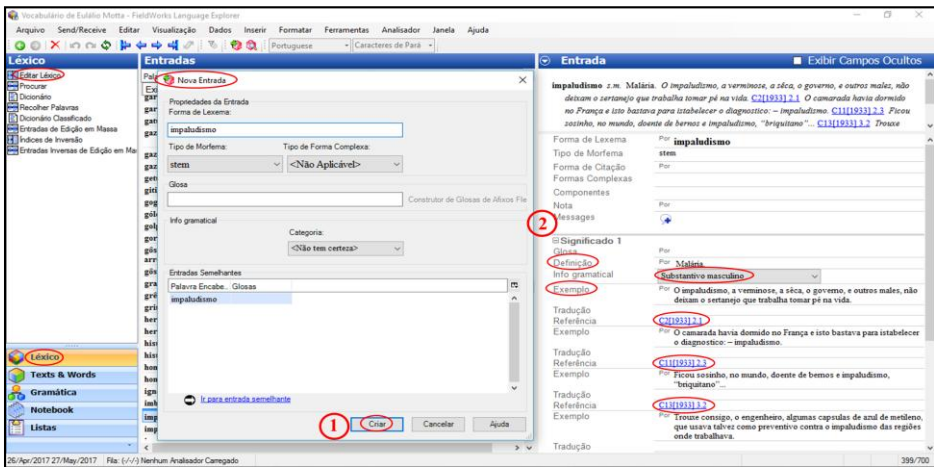


Figura 20 – Adição de nova entrada na área *Léxico* do FLEX
Fonte: Elaborada pela pesquisadora.

Após adicionar a nova entrada no painel central, o programa abre uma tela no painel de detalhes, onde é possível inserir a definição, a classificação gramatical, as variantes (quando houver), as abonações, além de som e imagem para ilustrar a entrada lexical. A busca pelos exemplos no *corpus* pode ser feita automaticamente, sendo a referência gerada a partir do código do arquivo pré-estabelecido na edição dos metadados. Tudo é feito com a aprovação do usuário, sendo possível fazer alterações.

A segunda forma adotada para acrescentar as entradas ao *Vocabulário de Eulálio Motta* foi pela aba *Glosa em Texts & Words*, como se vê na Figura 21. Nela, o *corpus* é fragmentado e numerado por sentença, abrindo uma caixa de diálogo que permite definir todos os itens lexicais, juntamente com a classificação gramatical, e adicioná-las como entrada. Além disso, é possível incluir mais de uma acepção para a mesma entrada ou criar entradas diferentes para as palavras homógrafas.

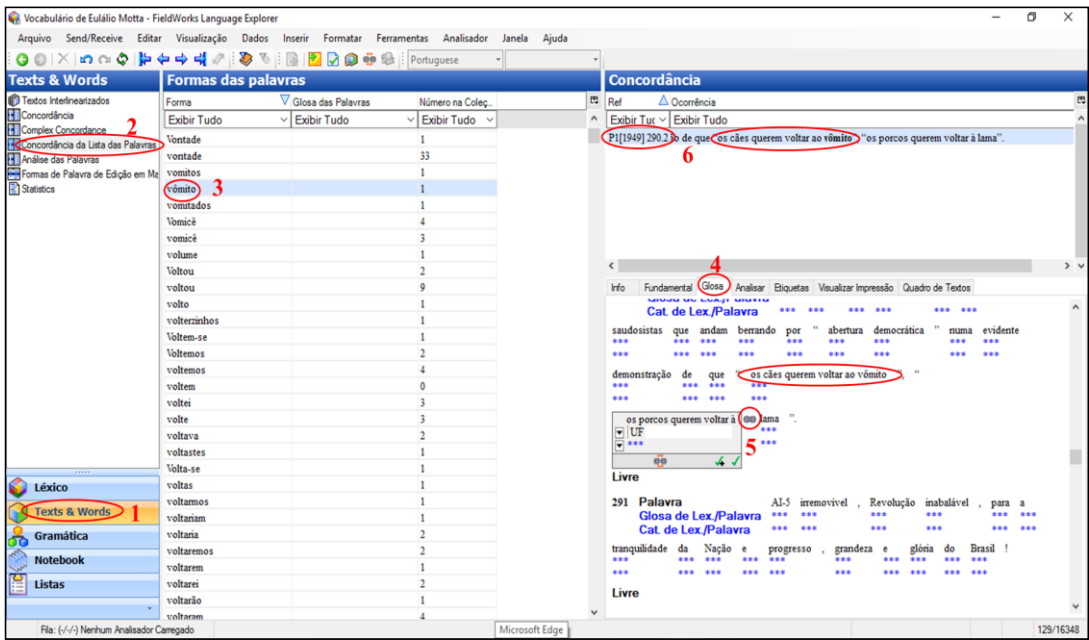


Figura 21 – Seleção e adição de nova entrada na área *Texts & Words* do FLEx
Fonte: Elaborada pela pesquisadora.

No painel central, visualizam-se todas as formas de palavras encontradas no *corpus* com o respectivo número de ocorrências. Além da praticidade em poder escolher o lema diretamente do texto, essa aba dispõe de um link que permite unir as palavras. Esta função é imprescindível para elencar as entradas de lexias compostas e complexas.

Na área *Gramática*, ajustou-se, de acordo com as demandas do *corpus*, o modelo de classificação proposto pelo programa. Nesse sentido, diferenciaram-se substantivos masculinos (s.m.) de femininos (s.f.) e os verbos transitivos (v.t.) dos intransitivos (v.i.). Acrescentaram-se também advérbios (adv.), adjetivos (adj.) e unidades fraseológicas (UF). Em *Listas*, traduziu-se a classificação dos tipos de variantes para que pudessem ser registradas no vocabulário.

Finalizada a edição das entradas do vocabulário, a visualização do trabalho fica disponível na área *Léxico*, na aba *Dicionário* (cf. Figura 22). Em *Ferramentas*, no menu principal, o usuário pode configurar o que e como será exibido. O *FLEX* organiza o trabalho lexicográfico em formato PDF e possibilita a impressão. Basta acessar a opção *Imprimir* na aba *Arquivo*, como se observa na Figura 22:

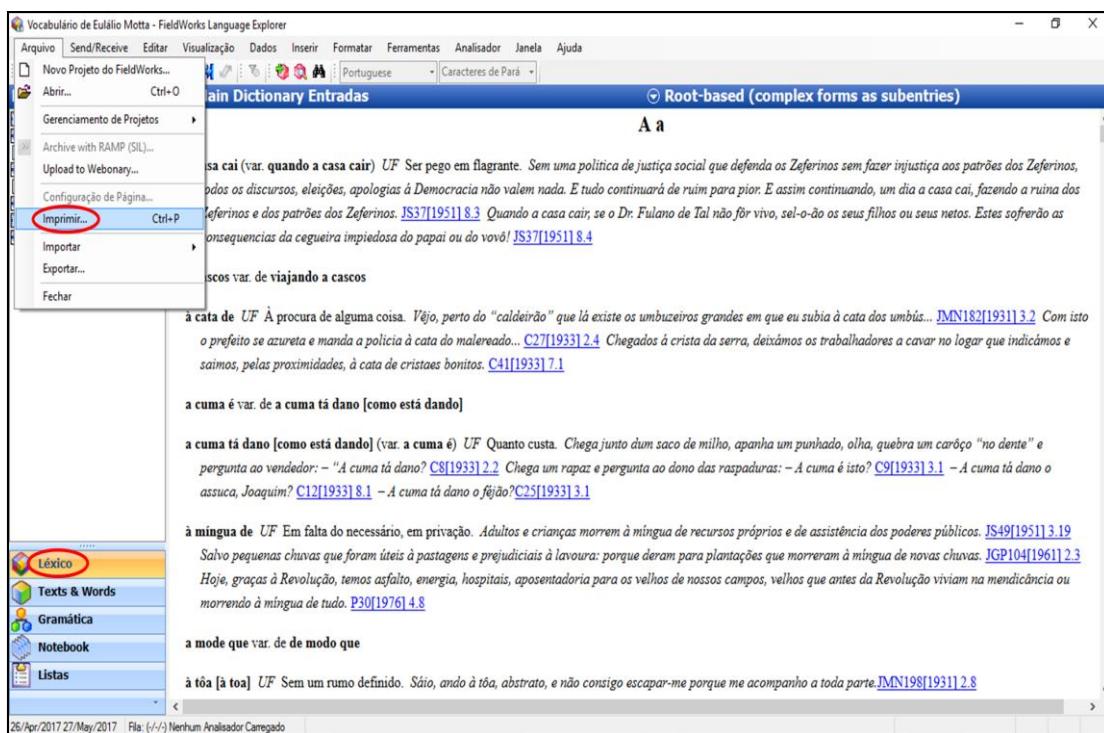


Figura 22 – Janela de impressão do *FLEX*

Fonte: Elaborada pela pesquisadora.

É possível, também, exportar o trabalho final para outros programas que fazem a edição em formato de dicionário e possibilitam disponibilizar *online*. Para o *Vocabulário de Eulálio Motta*, optou-se por exportá-lo para o programa *Pathway*¹¹, que é integrado a interface da *SIL FieldWorks*. O *Patway* gera o arquivo editável e transfere para o *LibreOffice Writer*, onde é possível alterar a formatação do estilo de dicionário padrão, como se vê na Figura 23.

¹¹ *Pathway* disponível em: <<http://pathway.sil.org/>>. Acesso em: 29 maio 2017.

A a

a casa cai (var. **quando a casa cair**) UF Ser pego em flagrante. Sem uma política de justiça social que defenda os Zeferinos sem fazer injustiça aos patrões dos Zeferinos, todos os discursos, eleições, apologias à Democracia não valem nada. E tudo continuará de ruim para pior. E assim continuando, um dia a casa cai, fazendo a ruína dos Zeferinos e dos patrões dos Zeferinos. JS37[1951] 8.3 Quando a casa cair, se o Dr. Fulano de Tal não for vivo, sel-o-ão os seus filhos ou seus netos. Estes sofrerão as consequências da cegueira impiedosa do papai ou do vovô! JS37[1951] 8.4

a cascos UF Longa caminhada de animais de um município a outro para ser abatido. E principalmente: as botadas gordas de Mundo Novo, Rui Barbosa, Monte Alegre, Macajuba, enfim: de toda esta zona de intensa produção; botadas que continuam viajando a cascos para o abate em Salvador! P16[1971] 4.4 A cascos, perdendo centenas de arrobas em cada boiada! A cascos, como na era dos carros de bois! A cascos, nesta era do asfalto! P16[1971] 5.1 "As boiadas seriam transportadas como as de Minas, em poucas horas em vez de muitos dias, em veículos modernos em vez de penosas, longas e tristes caminhadas a cascos. P16[1971] 8.2

à cata de UF À procura de alguma coisa. Vêjo,

Joaquim? C12[1933] 8.1 – A cuma tá dano o fêjão? C25[1933] 3.1

à mingua de UF Em falta do necessário, em privação. Adultos e crianças morrem à mingua de recursos próprios e de assistência dos poderes públicos. JS49[1951] 3.19 Salvo pequenas chuvas que foram úteis à pastagens e prejudiciais à lavoura: porque deram para plantações que morreram à mingua de novas chuvas. JGP104[1961] 2.3 Hoje, graças à Revolução, temos asfalto, energia, hospitais, aposentadoria para os velhos de nossos campos, velhos que antes da Revolução viviam na mendicância ou morrendo à mingua de tudo. P30[1976] 4.8

a mode que var. de **de modo que**

à tóa [à toa] UF Sem um rumo definido. São, ando à tóa, abstrato, e não consigo escapar-me porque me acompanho a toda parte. JMN198[1931] 2.8

abafa a banca [abafar a banca] UF Ser vitorioso em qualquer empreitada. Volta-se a verificar que a raça indubrasil é o grande coringa do mundo bovino. Como qualquer raça o mestiço de indubrasil abafa a banca. P37[1978] 4.5

abaixadora s.f. Subestação de energia. Roberto Santos, corresponda a esta vitória da Arena

Figura 23 – Exemplo do *Vocabulário de Eulálio Motta*

Fonte: Elaborada pela pesquisadora.

4 CONSIDERAÇÕES FINAIS

O *Vocabulário de Eulálio Motta* busca ser representativo de um universo de discurso (BARBOSA, 2001) de Eulálio Motta e visa como produto final “uma lista exaustiva das ocorrências que figuram num *corpus*” (DUBOIS et al., 1993 [1978], p. 613). Classifica-se, de acordo com os critérios de Haensch (1982), como um estudo de caráter linguístico, pois define palavras designativas (voltadas a referentes extralinguísticos), além de registrar a classe gramatical e o gênero das unidades lexicais lematizadas. Baseia-se no sistema linguístico individual de um autor, sendo todos os exemplos extraídos dos textos-fonte de estudo, que se configuram como abonações. É uma obra de constituição monolíngue por contemplar apenas unidades léxicas da língua portuguesa. Apresenta uma seleção lexical parcial e seletiva, pois registra uma parcela do léxico da língua, sincrônica e de caráter descritivo, que permite identificar as marcas de uso diafásica (pela riqueza da linguagem formal e informal utilizada), diastrática (por permitir o acesso a um vocabulário específico de uma comunidade linguística pouco documentada) e diatópica (por traduzir

os valores e as crenças do povo sertanejo de uma determinada região). As entradas foram ordenadas semasiologicamente, em ordem alfabética, partindo dos significantes para os significados.

O trabalho lexicográfico empreendido evidenciou as marcas linguísticas do escritor Eulálio Motta e de seu entorno sociocultural. A seleção lexical empreendida por ele, na elaboração de seus textos, retrata a sua visão de mundo, a sua formação intelectual, ao mesmo tempo em que documenta a cultura e a história de um grupo social em uma determinada época. As unidades lexicais que compõem a macroestrutura do *Vocabulário de Eulálio Motta* remetem às histórias locais, expressões populares, crenças, festas populares, a questões políticas, sociais e religiosas, ao cotidiano do homem sertanejo do município de Mundo Novo e regiões circunvizinhas.

Como resultado principal, obteve-se o *Vocabulário de Eulálio Motta*, composto por 700 entradas, que foram classificadas em: substantivo feminino, substantivo masculino, verbo transitivo, verbo intransitivo, adjetivo, advérbio e unidade fraseológica (BARREIROS; TELLES, 2017), sendo 505 lexias simples, 61 lexias compostas e 134 lexias complexas.

Este trabalho teve por finalidade contribuir para a preservação de costumes e valores culturais do homem sertanejo, expresso no seu uso da língua, e corroborar a importância de se preservar, através de textos literários, a cultura, a língua e a história local do povo sertanejo do Semiárido Baiano. A partir dos dados obtidos, vários aspectos podem ser explorados em futuras pesquisas, relacionadas à formação histórica da Língua Portuguesa e à variação linguística.

REFERÊNCIAS

- ANTHONY, Laurence. AntConc (Versão 3.4.3) [Software de Computador]. Tóquio, Japão: Universidade de Waseda. 2014. Disponível em: <<http://www.laurenceanthony.net/>>. Acesso em: 16 fev. 2017.
- BARBOSA, Maria Aparecida. Dicionário, vocabulário, glossário: concepções. In: ALVES, Ieda Maria. (Org.). *A constituição da normalização terminológica no Brasil*. 2. ed. São Paulo: FFLCH/CITRAT, 2001, p. 23-45.
- BARREIROS, Liliane L. S. *O vocabulário de Eulálio Motta*. 359f. Tese (Doutorado em Letras e Linguística) – Instituto de Letras, Universidade Federal da Bahia, Salvador, 2017.
- BARREIROS, Liliane L. S. *Bahia Humorística: causos sertanejos de Eulálio Motta*. Feira de Santana-BA: UEFS Editora, 2016.
- BARREIROS, Liliane L. S.; TELLES, Célia M. As unidades fraseológicas no Vocabulário de Eulálio Motta. *Revista Virtual de Estudos da Linguagem*, v. 15, p. 17-36, 2017.
- BARREIROS, Patrício N. *O pasquineiro da roça: a hiperedição dos panfletos de Eulálio Motta*. Feira de Santana-BA: UEFS Editora, 2015.

- BARREIROS, Liliâne Lemos Santana; BARREIROS, Patrício Nunes. O vocabulário da Ditadura Militar nos panfletos de Eulálio Motta. *Linguística e Filologia Portuguesa* (USP), São Paulo, v. 17, n. 2, p. 385-420, jul./dez., 2015.
- BARREIROS, Patrício Nunes. Eulálio Motta: um panfletário no sertão da Bahia. *Revista do Instituto de Estudos Brasileiros*, n. 67, São Paulo, p. 57-80, 2017.
- BARREIROS, Patrício Nunes. Por uma abordagem da História Cultural das práticas de escrita na edição de textos. *Alea: Estudos Neolatinos* (Impresso), v. 19, p. 389-414, 2017.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Manole, 2004.
- BIDERMAN, Maria Tereza C. *Teoria lingüística: teoria lexical e lingüística computacional*. 2 ed. São Paulo: Martins Fontes, 2001.
- BIDERMAN, Maria Tereza C. A ciência da Lexicografia. *Alfa Revista de Linguística*. São Paulo, n. 28 (supl.), p. 1-26, 1984.
- BIDERMAN, Maria Tereza C. *Teoria lingüística: lingüística quantitativa e computacional*. Rio de Janeiro: Livros Técnicos e Científicos, 1978.
- BIDERMAN, M. T. C. *A Categoria do Gênero* (nas línguas românicas). Tese (Livre-Docência). São Paulo: USP, 1974.
- BIDERMAN, M. T. C. *Análise Computacional de Fernando Pessoa* (Ensaio de Estatística Léxica). Tese (Doutorado em Filologia e Língua Portuguesa). Faculdade de Filosofia, Letras e Ciências Humanas/USP, São Paulo, 1969.
- DUBOIS, Jean et al. *Dicionário de Linguística*. 9. ed. Tradução de Frederico Pessoa de Barros et al. São Paulo: Cultrix, 1993 [1978].
- HAENSCH, Günther. Tipología de las obras lexicográficas. In: HAENSCH, G. et al. *La lexicografía: de la lingüística teórica a la lexicografía práctica*. Madrid: Gredos, 1982, p. 95-187.
- OTHERO, Gabriel de A.; MENUZZI, Sérgio de M. *Linguística computacional: teoria e prática*. São Paulo: Parábola, 2005.
- POTTIER, Bernard. *Lingüística general: teoria y descripción*. Trad. Maria Victoria Cantalina. Madrid: Gredos, 1977.

Recebido em: 17/05/2017
Aprovado em: 25/08/2017
Publicado em: 01/12/2017