

Resumo: O objetivo deste texto é descrever os recursos oferecidos pelo *WebCorp* (*The Web as a Corpus*) e mostrar como podem ser utilizados para a extração de informação linguística que auxiliam os tradutores a encontrarem soluções no decorrer do processo tradutório e também a aprimorar sua competência tradutória na medida em que adquire maior conhecimento lingüístico, extralingüístico e instrumental.

Introdução

O *WebCorp* (*The Web as a corpus*) é um conjunto de ferramentas que permite o acesso a *World Wide Web* como um *corpus* voltado à busca de informação linguística. Está disponível livremente em www.webcorp.org.uk e foi criado e é mantido pela Research and Development Unit for English Studies (RDUES) na School of English, Birmingham City University.

Descrevemos, a seguir, o *WebCorp*, apresentando os recursos que oferece e mostrando sua aplicação na tradução, principalmente, na busca de soluções de tradução e, em consequência, no aprimoramento contínuo das subcompetências tradutórias (Hurtado Albir, 2001 e 2005) que conformam a competência tradutória. De forma resumida, apresentamos as seis subcompetências propostas pela autora:

- a) Bilíngue: conhecimento lingüístico sobre as línguas de trabalho do tradutor;
- b) Extralinguística: conhecimento de mundo, incluindo as culturas das línguas de trabalho;
- c) Instrumental: uso de fontes e recursos de informação e conhecimento sobre a profissão (valores cobrados, associações de tradutores, etc.);
- d) Conhecimentos de tradução: conhecimentos sobre os princípios que guiam a tradução como processos, métodos e estratégias tradutórias;
- e) Estratégica: considerada a competência central que rege as demais e se relaciona aos procedimentos operacionais como, por exemplo, definição do tempo para realizar uma tradução, constituição de equipe para realização de determinada tradução, etc.
- f) Componentes psicofisiológicos: referem-se à atenção, memória, curiosidade intelectual, capacidade de síntese, etc.

Ao longo do texto, procuraremos associar as informações extraídas com os recursos do *Webcorp* às subcompetências tradutórias.

[□] Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

Embora o foco do texto esteja voltado à tradução, ressaltamos que os recursos oferecidos pelo *WebCorp* também podem ser utilizados para outras finalidades como, por exemplo, o ensino e aprendizagem de línguas estrangeiras e línguas para propósitos específicos.

Recursos oferecidos pelo *WebCorp* e sua aplicação na tradução

O *WebCorp* oferece basicamente dois tipos de busca: simples e avançada. A busca simples inclui os campos: *Keyword* (palavra-chave ou palavra de busca), seleção do buscador, *sensitive case* (seleção de palavras iniciadas por maiúsculas ou minúsculas), *span* (extensão do contexto da palavra-chave, isto é, se queremos 4, 5, 6 ou mais palavras à esquerda e à direita da palavra de busca), e a língua sobre a qual queremos obter informações.

Por sua vez, a busca avançada (*advanced options*) oferece outros filtros além dos incluídos na busca simples: domínio na *web* (.br, .es, .uk, .it, etc.), *sites* populares e área ou tema (*Word filter*) relacionado à busca. Por exemplo, para buscar informações sobre o uso do termo *lúpus* somente no português do Brasil podemos inserir o domínio *.br*, excluindo, assim, a busca de textos escritos nos demais países de língua portuguesa. Mostramos a tela da busca avançada exemplificando com o termo *lúpus* associado a *reumatismo* (filtro) e, a seguir, alguns dos resultados obtidos com essa busca.

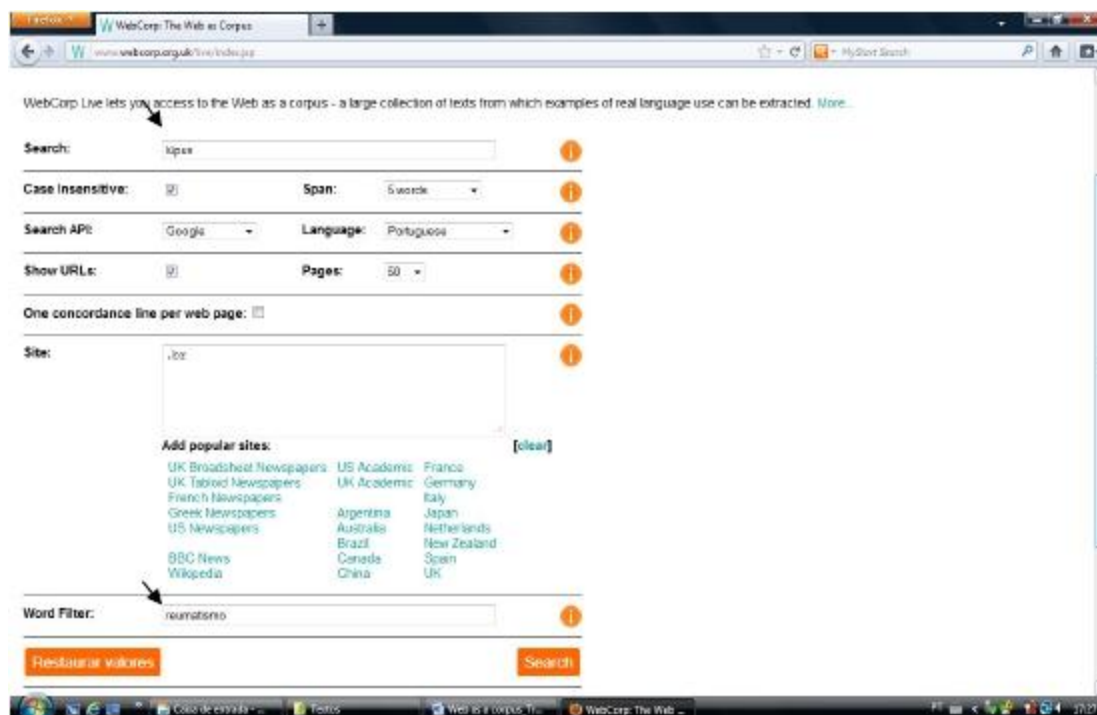


Fig. 1 – Busca avançada

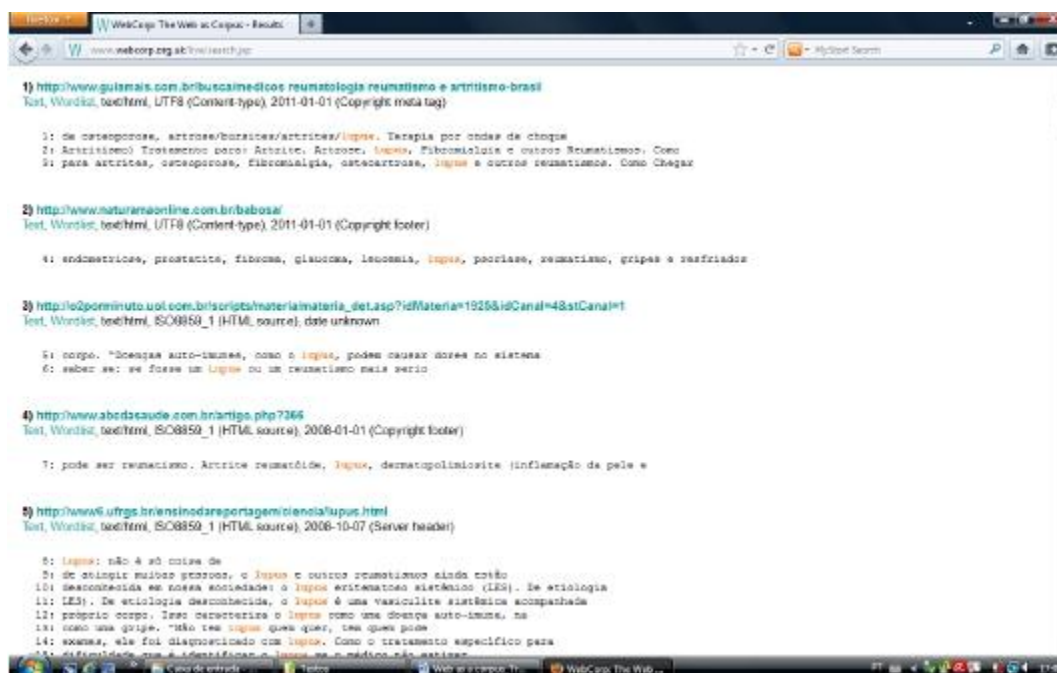


Fig. 2 – Resultados da busca avançada para o termo *lupus* e filtro *reumatismo*

Na tela acima, cada uma das linhas é uma concordância, isto é, o contexto de uso da palavra buscada. A extensão do contexto é definida na tela inicial de busca, conforme já dissemos anteriormente. No exemplo apresentado, optamos por 5 palavras à esquerda e 5 à direita da palavra de busca.

Mas para que podem servir as concordâncias de determinada palavra ou termo para um tradutor? Várias podem ser as respostas. Entre elas, podemos dizer que as concordâncias podem servir para:

- Identificar termos sintagmáticos, como *lupus eritematoso sistêmico* (5, linha 10 acima);
- Identificar a sigla do termo *lupus eritematoso sistêmico*: LES (5, linha 10);
- Obter informações sobre a palavra ou termo buscado. No exemplo de *lupus*, é possível saber que é uma doença autoimune (3, linha 5) e que faz parte de um grupo de doenças reumatóides como artrite e artrose (1, linha 3);
- Reconhecer formas de dizer típicas da Medicina; p. ex., *ele foi diagnosticado com lupus* (5, linha 14) e não *foi identificado com lupus*, por exemplo.

Tais informações ajudam a aprimorar as subcompetência linguística ou bilíngue (a, b e d), extralinguística e instrumental (c).

Além de oferecer os contextos da palavra-chave, se clicarmos sobre ela, abre-se uma nova janela com o texto completo, tal como vemos a seguir. A palavra de busca aparece

ressaltada em amarelo. Há ainda a fonte do texto, dado importante para garantir a confiabilidade da informação obtida¹.

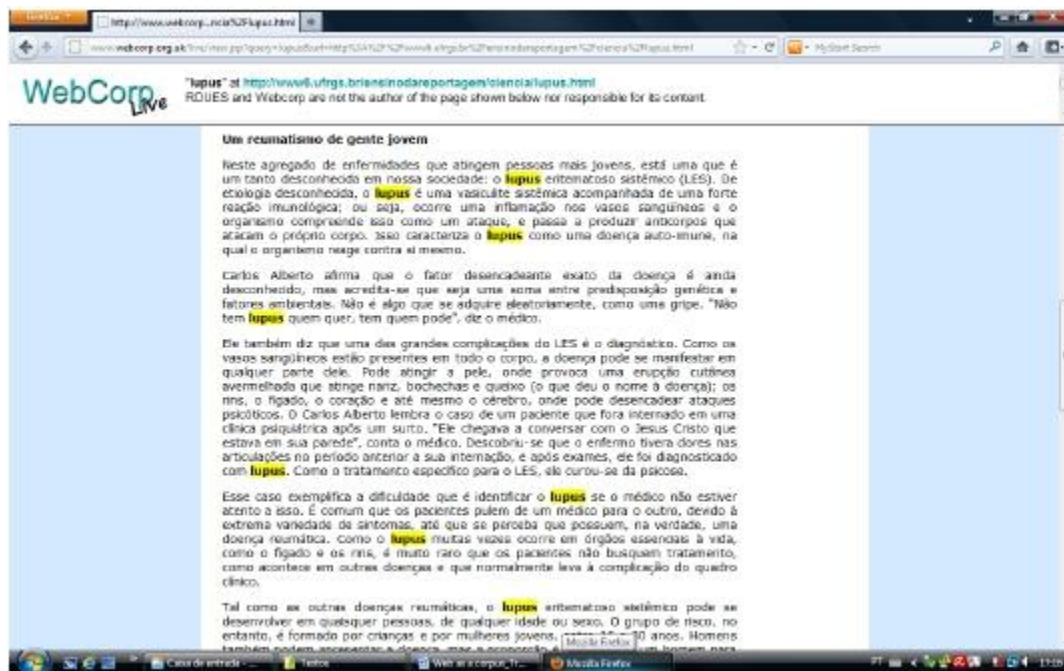


Fig. 3 – Texto completo com palavra de busca destacada.

Ter os contextos completos e o texto integral da palavra buscada pode servir ao tradutor para complementar as informações encontradas nas concordâncias. Permite principalmente observar:

a) o funcionamento da expressão de busca em seu contexto de uso, isto é, em que situação comunicativa é usada. No exemplo acima, podemos ver que é um texto de caráter menos especializado, voltado a um público não especialista, o que pode ser identificado pelo uso de marcadores do discurso como *isto é*, indicador de explicações ou explicitações, e por frases mais informais como “Não tem lúpus quem quer, tem quem pode”;

b) a estrutura sintática em que a expressão é utilizada, aspecto importante para o tradutor, pois esta estrutura pode mudar de uma língua a outra. Por exemplo: “ele foi diagnosticado com lúpus”; “identificar lúpus”, “como o lúpus muitas vezes ocorre em órgãos essenciais...”. Tais exemplos permitem ver que os verbos coocorrentes do termo *lúpus* são: *diagnosticar* em estrutura passiva (muito freqüente em artigos científicos em português), *identificar* e *ocorrer*;

¹ Para mais informações sobre critérios para seleção de informações na Web, ver Marins e Krieger (2005), disponível em http://www6.ufrgs.br/termisul/biblioteca/artigos/artigo_ABECAN_2005_MARINS_KRIEGER.pdf

c) obter definições sobre a expressão buscada: *lúpus* “é uma vasculite sistêmica acompanhada de uma forte reação imunológica”;

d) obter outras informações úteis para que o tradutor amplie e aprofunde seu conhecimento sobre o tema que está traduzindo: no caso do texto acima, sabe-se que há incerteza sobre a causa do *lúpus* e a dificuldade de se fazer seu diagnóstico.

Tais dados servem para aprimorar as subcompetências linguística (b e c) e extralinguística (a, b e d). Além de oferecer indícios para tomada de decisões em relação aos procedimentos e estratégias de tradução (subcompetência sobre os conhecimentos de tradução), na medida em que são obtidas informações úteis para encontrar soluções de tradução (compreensão sobre determinado termo e a busca de seu(s) equivalente(s), por exemplo).

Também é possível gerar a lista de palavras (no menu, *Wordlist Tool*) de cada um dos textos, clicando em *wordlist*, logo abaixo da fonte do texto.

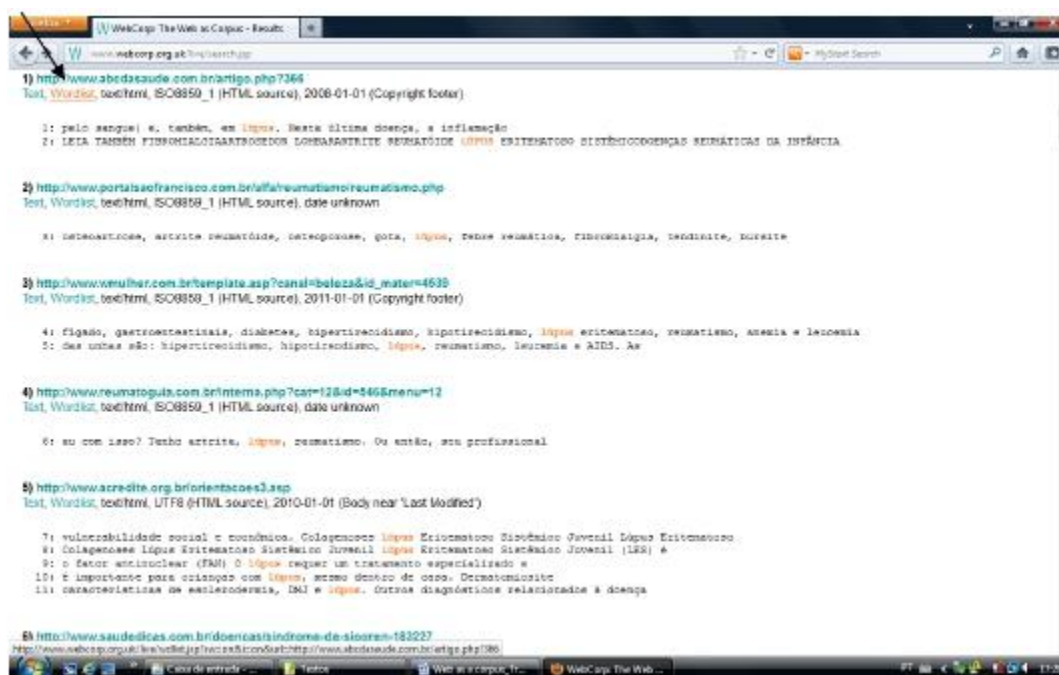


Fig.4 – Geração de *Wordlist*

Uma amostra desta lista para o primeiro texto obtido para o termo *lúpus* pode ser vista abaixo. Nela, vemos que há palavras gramaticais (*que, é, mais*, etc.), mas também palavras lexicais que são termos relacionados ao item buscado (*artrite, reumatismo, doença, saúde*):

Word	Frequency
que	27
e	23
artrite	16
reumatismo	14
mas	14
doença	11
doenças	11
com	11
saúde	10
não	9
em	9
são	8
ser	8
abc	8
um	8
mas	7
do	7
inflamação	6
esta	6
como	6
câncer	6
transito	6

Fig.5 – Lista de palavras do texto

Essa lista pode mostrar ao tradutor:

- o possível tema do texto, isto é, ao observar as palavras lexicais (*artrite*, *reumatismo*) mais frequentes, pode-se ter pistas sobre a temática tratada no texto;
- os possíveis termos: *artrite*, *reumatismo*, *inflamação*, *câncer*;
- os temas desconhecidos do tradutor e sobre os quais deverá buscar e obter novos conhecimentos (conhecimento extralingüístico) para poder fazer uma tradução de melhor qualidade.

As informações obtidas reforçam as subcompetências *bílingue*, extralinguística e instrumental.

Pode-se ainda baixar o texto em formato txt, clicando em *text*, logo abaixo da fonte do texto, tal como vemos na tela a seguir. O resultado obtido pode ser visto na figura 7 que traz uma amostra de um texto.

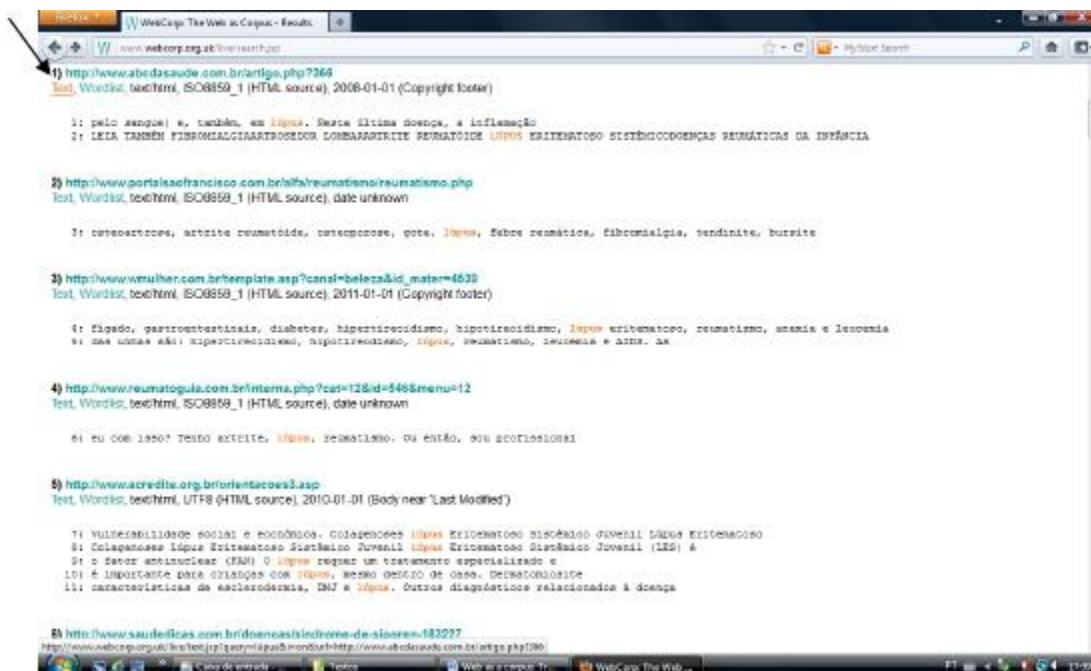


Fig 6 – Geração de texto em formato txt

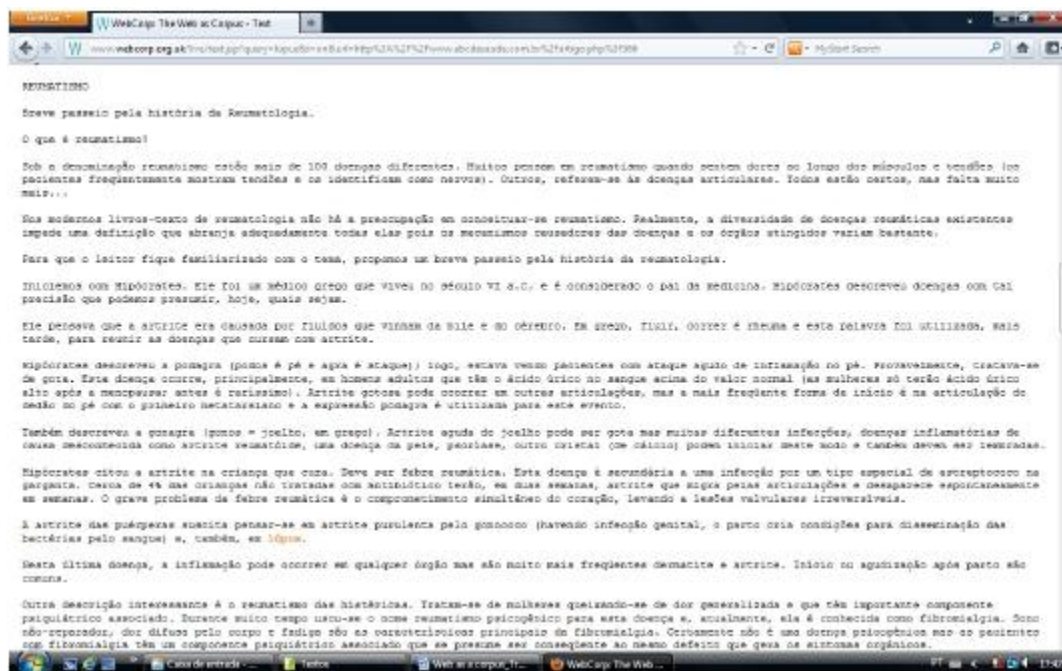


Fig. 7 – Texto em formato txt

Baixar o texto completo em txt pode servir para construir *corpora* textuais da temática sobre a qual se está traduzindo. Isto significa dizer que é possível selecionar um conjunto de textos neste formato e salvá-los em arquivos ou pastas específicas. Posteriormente, estes textos poderão ser processados por outras ferramentas de extração

linguística - AntConc² e Corpógrafo³, por exemplo. Essas ferramentas de extração linguística permitem gerar listas de palavras, concordâncias, *n-gramas* e *clusters* que também auxiliam no processo de tradução, tal como estamos mostrando com o WebCorp.

Outra ferramenta disponível é a geração de *n-gramas*, isto é, a lista de uma palavra (unigrama), duas palavras (bigrama) até cinco palavras (pentagrama). Na tela abaixo, incluímos a página do Dr. Drauzio Varela⁴ e selecionamos trigramas:

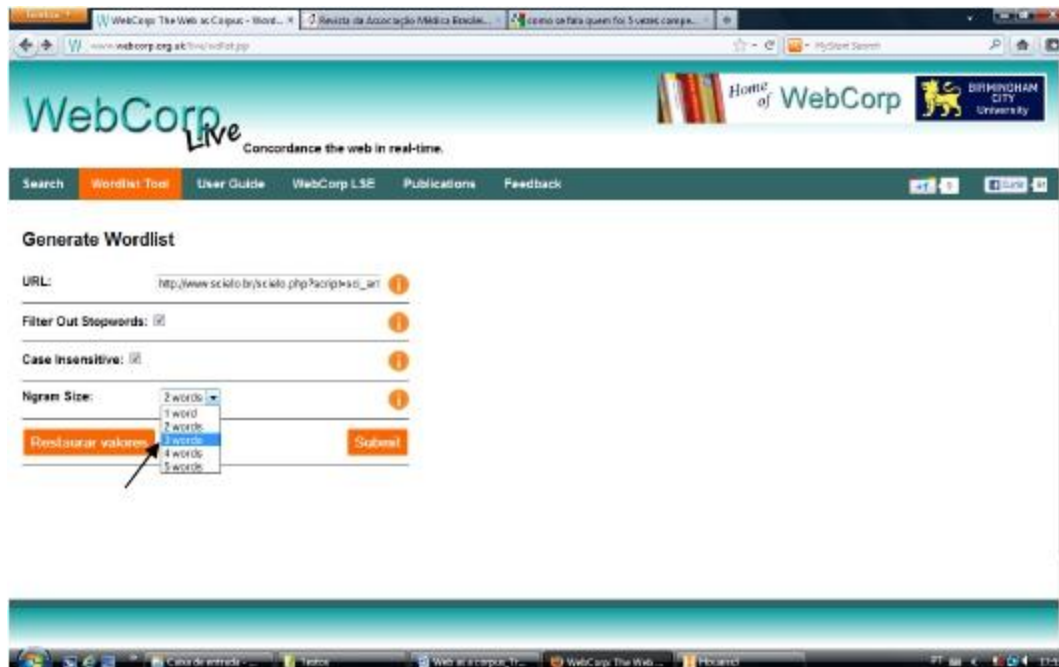


Fig 8 – Geração de trigramas

Como resultados da busca, temos abaixo termos e expressões relacionadas ao lúpus: *lúpus eritematoso sistêmico*, *doenças da pele*, *exposição ao sol*, *asa de borboleta*, *manchas na pele*, *agentes patogênicos*, etc. Além dessa informação de caráter mais terminológico, clicando em cima da expressão, abre-se o texto completo, tal como mostramos anteriormente nas concordâncias da palavra de busca.

² Disponível em <http://www.antlab.sci.waseda.ac.jp/software.html>

³ Disponível em: <http://www.linguateca.pt/corpografo/>

⁴ <http://drauziovarella.com.br/doencas-e-sintomas/lupus-eritematoso-sistmico/>

Ngram	Frequency
lupus eritematoso sistêmico	3
exposição ao sol	2
doenças de pele	2
a exposição ao	2
usa de borbulha	2
recomendações na boca	2
e no nariz	2
manchas na pele	2
do lupus eritematoso	2
afetado os mais	1
los estudos recentes	1
adante no entanto	1
de agentes patogênicos	1
modernos apresentam menos	1
e usar protetores	1
tem positivo tratamento	1
recomendações pacientes com	1
humana apoio lemmes	1
órgão afetado os	1
da doença os	1
depressão protata garbrie	1

Fig. 9 – Resultados da geração de trigramas

Essa busca oferece resultados que auxiliam na subcompetências linguística, extralinguística e instrumental, tal como vimos para os resultados obtidos para as outras ferramentas do *Webcorp*.

Para finalizar, cabe destacar que, embora a maioria dos resultados gerados pelas diferentes ferramentas propiciem uma melhoria ou aprofundamento das competências bilíngue, extralinguística e instrumental, eles igualmente contribuem para as demais subcompetências tradutórias. Sendo assim, a subcompetência sobre conhecimentos de tradução, a subcompetência estratégica e os componentes psicofisiológicos serão ativados em cada um dos procedimentos realizados. Por exemplo, ao gerar os resultados dos trigramas, se ativará a subcompetência sobre conhecimento de tradução, uma vez que é preciso saber identificar quais das unidades geradas podem ser consideradas uma unidade de tradução e se esta é o equivalente adequado para a unidade da língua de partida ou ainda saber como ela é utilizada na língua de chegada. Para tomar tais decisões, o tradutor necessariamente ativará a competência estratégica, definindo os procedimentos e estratégias que melhor se adequam ao texto a ser traduzido. Além disso, mobilizará os componentes psicofisiológicos, pois deve buscar em sua memória outros conhecimentos já adquiridos sobre o tema do texto a ser traduzido, deve prestar atenção e analisar com cuidado os dados obtidos para tomar as decisões adequadas, entre outros aspectos relacionados a estes componentes.

Conclusões

Com essa breve apresentação, foi possível mostrar as ferramentas e os recursos oferecidos pelo *WebCorp* e estabelecer sua relação com as subcompetências tradutórias. É importante ressaltar que tais recursos servem de subsídios para o processo tradutório e oferecem, sobretudo, informações linguísticas contextualizadas, refletindo seus usos em textos reais. Possibilitam, portanto, ao tradutor observar os fenômenos lingüísticos das suas línguas de trabalho e fazer as melhores escolhas para traduzir um texto que seja correto linguisticamente - com usos de colocações, de expressões, etc. - na língua de chegada, e adequado do ponto de vista da temática tratada – utilização de termos, variação linguística, adequação ao nível de especialização, etc. Em suma, utilizar os recursos oriundos da Linguística de *Corpus* - como são as ferramentas do *WebCorp* – auxilia o tradutor a melhorar a sua competência tradutória e, conseqüentemente, a qualidade do texto traduzido.

Referências bibliográficas

- BEVILACQUA, C.R. *Traducción, Terminología y Fraseología Especializada: relaciones necesarias para la adquisición de la competencia traductora*. Colegio de Traductores Públicos de la Ciudad de Buenos Aires. (no prelo).
- HURTADO ALBIR, A. A aquisição da competência tradutória. Aspectos teóricos e didáticos. In: PAGANO, A; MAGALHÃES, C., ALVES, F. (org.). *Competência em tradução. Cognição e discurso*. Belo Horizonte: Editora UFMG, 2005, p. 19-57.
- HURTADO ALBIR, A. *Traducción y traductología. Introducción a la traductología*. Madrid: Cátedra, 2001.