

# GLOSSANET: ELABORAÇÃO DE MATERIAIS DE LÍNGUA ESPANHOLA

Cristiane Magalhães Bissaco<sup>1</sup>

**RESUMO:** O Glossanet tem sido apontado como um software útil no levantamento de linhas de concordância de diversas línguas. O uso de tais linhas, bem como a utilização de corpus de língua autêntica, trazem uma contribuição bastante significativa para o processo de ensino-aprendizagem de língua estrangeira. Assim, a pesquisa aqui retratada busca mostrar quão relevante é sua utilização para o ensino específico de língua espanhola para o curso de Graduação em Administração de Empresas com habilitação em Comércio Exterior.

**Palavras-chave:** Lingüística de Corpus, linhas de concordância, processo de ensino-aprendizagem de língua estrangeira.

**ABSTRACT:** *Glossanet has been considered a useful software for getting concordance lines in different languages. The use of such lines and authentic language corpora have a significant contribution for the foreign language teaching and learning process. Therefore, this paper presents a research which aimed at showing the relevance of using concordances in teaching Spanish in an undergraduate Business course with emphasis on Foreign Commerce.*

**Keywords:** *Corpus Linguistics, concordance lines, foreign language teaching and learning process.*

## 1. Introdução

A preocupação com o processo ensino-aprendizagem de língua estrangeira – espanhol, tem me motivado como pesquisadora. Atualmente, como professora de um Centro Universitário do interior paulista, trabalho com o Curso de Administração de Empresas com habilitação em Comércio Exterior, ministrando aulas de Espanhol Instrumental, o que me levou à necessidade de elaborar materiais específicos ao curso, já que no Brasil não disponibilizamos de muitas opções.

---

<sup>1</sup> A autora é professora de língua espanhola do Centro Universitário Toledo de Araçatuba, São Paulo. Especialista em Lingüística e Língua Portuguesa pela UNESP – São José do Rio Preto, mestre em Lingüística Aplicada pela PUC-SP e doutoranda no programa de Lingüística Aplicada pela mesma instituição.

Com o propósito de aplicar material autêntico em sala de aula, em vez de utilizar os didáticos que retratam uma ficção no ensino de segunda língua, comecei a levantar meu próprio material a partir de textos da área de economia disponibilizados em sites internacionais. O termo autêntico deve ser entendido como a ocorrência natural da linguagem, ou seja, linguagem que não tenha sido criada especificamente para situações de ensino, tais como a transcrição de uma conversa entre falantes nativos, ou de um texto de jornal.

Portanto, tenho utilizado textos jornalísticos on-line, recorrendo a alguns jornais argentinos e espanhóis, entre eles La Nación e El Mundo, como também, à produção norte americana chamada CNN en Español. Entretanto, minha atuação com o material autêntico ainda parece deficiente. Assim, meu propósito é elaborar atividades para os alunos da graduação do curso de Administração de Empresas com habilitação em Comércio Exterior, que os auxiliem no que diz respeito a aquisição de vocabulário específico e de uma gramática contextualizada.

Basicamente, os materiais disponíveis no mercado fornecem as explicações gramaticais e de vocabulário sem permitir que o aprendiz tenha a oportunidade de descobrir como a língua se comporta, por meio da exploração de grandes números de instâncias retiradas de **corpus**, como é o caso das **concordâncias**.

Desse modo, este trabalho pretende sugerir que é possível desenvolver atividades que despertem no aprendiz a curiosidade para a investigação da linguagem com corpus e que contribuam para que o educando participe ativa e conscientemente da construção do seu conhecimento explorando padrões de linguagem.

O objetivo é mostrar a utilização pedagógica de uma abordagem baseada em **Linguística de Corpus**, a partir da frequência de palavras encontradas em textos jornalísticos e suas concordâncias, no desenvolvimento, aplicação e análise de atividades didáticas que estimulem a construção do conhecimento do aprendiz, no ensino de espanhol como língua estrangeira. Tais atividades terão como foco os padrões encontrados em um corpus composto por textos jornalísticos, previamente selecionados pelo software que será detalhado neste estudo: o **Glossa Net**.

## 2. Conceitos teóricos

Esta pesquisa apresenta como pressuposto teórico a **Linguística de Corpus**, já que se ocupa da coleta e da exploração de corpora, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para estudos de uma língua ou variedade lingüística, conforme destacado por Berber Sardinha (2004). Assim, corpus é sinônimo de corpo, conjunto de documentos e o conjunto desse corpo é determinado como corpora.

Para Sinclair (1991) o que possibilita a obtenção de bons resultados em pesquisas baseadas em um **corpus** são as decisões que o organizador deve tomar para selecionar seus textos e organizá-los. Uma das decisões que o organizador deve tomar está relacionada à forma da língua, oral ou escrita. Sinclair (1991) observa que, apesar das dificuldades para a coleta de dados, corpora montados a partir de exemplos da língua falada são mais significativos, o que não será o objetivo nesta pesquisa, pois pretendo me concentrar em textos jornalísticos selecionados pelo próprio programa Glossa Net.

Outra característica fundamental para a **criação de um corpus**, na opinião de Sinclair (1991), é a utilização de textos não fragmentados, pois apresenta-se mais adequada do que uma coleção de pequenas partes. Além disso, um corpus deve estar em constante desenvolvimento, primeiramente pelo fato da língua estar em contínua evolução, e em segundo lugar, um corpus grande pode, a critério do analista, tornar-se fonte de textos completos, para o caso de pesquisas de cunho mais específico.

Segundo Sánchez (1995, apud Berber Sardinha, 2000) um **corpus** é um conjunto de dados lingüísticos, sejam do uso escrito ou uso oral, sistematizados, com ampla extensão, de maneira que representem o uso lingüístico ou algum de seus âmbitos, como em minha pesquisa, por exemplo, o recorte é dado à área de economia; esses dispostos de tal modo que possam ser processados pelo computador, com a finalidade de propiciar resultados úteis para a descrição e análise.

Outro conceito apresentado nesta pesquisa é o da **concordância**. Segundo Berber Sardinha (2004), a concordância é uma listagem das ocorrências de um item específico, dispostas de tal modo que a palavra de busca aparece centralizada na página. A palavra de busca é acompanhada do seu contexto original, isto é, das palavras que ocorreram junto com ela no corpus.

### 3. Apresentação do software

A discussão sobre Lingüística de Corpus, uso de corpora e de linhas de concordância, fizeram-se necessárias para um melhor entendimento das funções que o software freeware nomeado Glossa Net é capaz de desempenhar.

Glossa Net é uma ferramenta com interface on-line específica para pesquisar corpora de jornais on-line, em diversas línguas, inclusive, português, que efetua concordâncias e localiza padrões on-line. Não há download do software, pois, a ferramenta funciona on-line. O cadastro e o uso são gratuitos.

O GLOSSANET traz em sua página inicial o tópico “Lingüística de corpus e busca de informação”. O GlossaNet é um motor e busca que dá acesso a edição do dia de mais de 100 jornais on-line em mais de 12 idiomas. Utiliza-se esse serviço para identificar informações na mídia escrita e por lingüistas a procura de estruturas sintáticas e enunciados atestados. Também o GlossaNet usa dicionários eletrônicos, por isso, os usuários podem refinar as suas requisições com informações morfológicas, sintáticas e semânticas. Estuda mais de 100 corpus diariamente atualizados em português, francês, holandês, inglês, italiano, grego, norueguês, espanhol,...

A primeira etapa é o cadastro do pesquisador, como dito anteriormente, a utilização do software é freeware, ou seja, não há custo para o usuário. O endereço de acesso é: <http://glossa.fltr.ucl.ac.be/indexbr.html>.

O GlossaNet foi desenvolvido pelo CENTAL (Centre de traitement automatique du langage) à Université de Louvain - Belgica. Tem uma versão em inglês e uma em francês para acessar. Você precisa digitar seu nome e sua senha e clicar uma vez em "login".

As categorias sintáticas que o programa oferece para localizar no corpus são: Adjetivo, substantivo, advérbio, preposição, verbo, artigo, pronome, conjunção coordenada e conjunção subordinada. Estas categorias são apresentadas em linhas de concordância.

Para adicionar uma nova tarefa clique uma vez em ADD TASK. O programa dá as opções de jornais disponíveis para você formar o corpus. Depois você seleciona os padrões a serem detectados. E por último a concordância. Digite seu "e-mail", pois é através dele que você

recebe as concordâncias ou em formato .html ou em formato .txt. A frequência dos resultados pode ser uma vez ao dia, uma vez por semana ou sempre que possível. Os caracteres a direita são 40 e a esquerda também 40.

#### 4. Utilizações do software em pesquisas acadêmicas

A partir da leitura de **Fairon & Courtois (2000)**, percebemos a utilização do software Glossa Net em um contexto de extensão de cobertura lexical dos dicionários eletrônicos em inglês.

Há no artigo uma discussão sobre os dicionários eletrônicos de uma maneira geral, tais como os de francês, espanhol, italiano, norueguês, português, grego e eslavo. Os autores afirmam que além da cobertura lexical, os dicionários trazem uma série de informações morfológicas, sintáticas e semânticas.

Assim, a utilização do **Glossa Net** é justificada pela extração de informações em pesquisas de estruturas lexicais e sintáticas. O processo pelo qual é realizada esta extração de informações se dá em um primeiro momento por um Corpus da Web que logo é convertido a um formato HTML e seus resultados são enviados ao e-mail do pesquisador.

Há no estudo de Fairon & Courtois (2000) uma preocupação com formação de palavras simples e compostas. Com base em 18 publicações do Washington Post, o que foi detectado em 1700 linhas de concordância é que 610 palavras tinham uma formação simples, enquanto 210 apresentavam uma formação composta.

Outro artigo bastante relevante em relação à aplicação do software em diferentes contextos de pesquisa é o de **Dister & Fairon (2000)**. Os autores se preocupam com as várias categorias de palavras não encontradas nos dicionários, como o caso de neologismos, nomes próprios e erros de ortografia e de pronúncia, com objetivo de ampliar dicionários.

A pesquisa é realizada por meio de um corpus dinâmico de sete jornais de Québec - Canadá (Le Soleil, La Presse, Le Nouvelliste, Le Droit, La Tribune, Le Quotidien e La Voix de l'Est) e de uma metodologia simplificada para a extração de palavras desconhecidas por

meio do software **Glossa Net**. Vale ressaltar que o conceito de corpus dinâmico é introduzido por Antoinette Renouff (1992, apud Dister & Fairon, 2000).

Os autores ressaltam que o programa é especializado na coleta de textos de jornais, em suas publicações on-line. São 80 jornais em 9 idiomas diferentes, como francês, inglês, italiano, norueguês, português, espanhol, grego, alemão e holandês. Os textos coletados são analisados com o auxílio de um programa chamado Unitex, um programa de análise de corpus que permite a aplicação de recursos lexicais em relação ao texto. Um desses recursos é dado pela aplicação do Glossa Net que forma linhas de concordância em formato HTML.

Dister & Fairon (2000) justificam que é por meio das linhas de concordância que o pesquisador em linguagem pode identificar formas-padrão, ou seja, se pode identificar a estrutura mais recorrente com aquele léxico, observando o que segue à direita e o que aparece à esquerda, com isso é possível por exemplo definir o uso de preposições que seguem substantivos. Os pesquisadores observam ainda uma ortografia reformada, o uso de apócoses, o uso de prefixos e sufixos, enfim, o uso de neologismos dentro do que os autores identificam como o “quebequismo”, uma língua derivada do francês, mas que possui particularidades marcantes e que precisam de estudos da área de linguagem.

**Rocha & Santos (2000)** realizam uma pesquisa com objetivo de elaborar um corpus de português europeu baseado nos extratos do jornal – O Público. Os autores mostram que o jornal foi fundado em 1990, e foi o primeiro a ser publicado simultaneamente em Lisboa e no Porto. Além disso, foi o primeiro jornal português com uma edição completa *online* (<http://www.publico.pt>), lançada em 1995. Este serviço fornece as últimas sete edições do jornal gratuitamente e disponibiliza vários serviços de informação exclusivamente na rede. Foi igualmente o primeiro (e até agora único) jornal português a publicar um livro de estilo (LEP). Finalmente, desde a sua fundação o *Público* tem fornecido material textual a vários grupos de I&D interessados no processamento da língua portuguesa, como se pode ver nos nomes dos corpora portugueses existentes no nosso catálogo: Natura/Público e BD-Público. Além disso, fornece o material português para o sistema de concordâncias por correio eletrônico **GlossaNet**, desenvolvido pelo LADL.

Apesar de existirem vários corpora de português de Portugal, não existia nenhum que se pudesse considerar de grandes dimensões. Assim, com o *Público*, os pesquisadores conseguiram um contrato para disseminar 100 milhões de palavras de texto corrido, retiradas das suas edições, em troca de uma pequena contrapartida financeira. A versão 1.0 do CETEMPúblico (Corpus de Extratos de Textos Eletrônicos MCT/Público), criada a 25 de Julho de 2000, contém cerca de 180 milhões de palavras distribuídas por 1.567.625 extratos, correspondentes a cerca de 1.500 edições diárias (algumas delas incompletas), quase inteiramente em português europeu. O corpus, além de distribuído gratuitamente em CD, organizado em vinte volumes, encontra-se incluído no serviço de acesso a corpora através da Internet (AC/DC).

**Fletcher (2005)** destaca a importância de se trabalhar com concordâncias em material da Web. O programa utilizado pelo autor é o KWICFinder (KF), cuja aplicação ajuda a perceber como a linguagem de estudantes e aprendizes é descrita e motivada em detalhe. O artigo discute o meio da Web em pesquisas sérias e propõe uma solução inicial. O autor descreve o futuro da Web para pesquisas de Corpus e sua aplicação.

Segundo, Fletcher (2005), o uso da Web para a composição de Corpus é muito relevante já que se pode notar a diversidade lingüística, entre outros fatores. Desse modo, o autor destaca o uso do **Glossa Net**, já que por meio do programa é possível perceber neologismos, estrangeirismos, nomes, enfim, um conjunto de novas palavras. O Glossa Net é, segundo o autor, um programa que monitora 100 jornais em 12 línguas, os dados publicados pelo programa são pesquisados pela estrutura ou pela formação da palavra. A limitação imposta, segundo, Fletcher (2005), é que infelizmente o material é coberto por somente alguns dias. O que o autor quer mostrar é que usando o KWICFinder (KF), o pesquisador tem acesso a um acúmulo de dados da Web, enquanto que usando o Glossa Net, os dados são vistos separadamente, seja por dia, ou por semana, de acordo com o interesse do pesquisador.

## 5. Objetivos desta pesquisa

A seleção do Corpus foi feita em Língua Espanhola em sete jornais. Os da Argentina: Clarin e La Nación; o do México: La Jornada e os da Espanha: ABC, El País, El Mundo, Gaceta.

O programa possui um registro de tarefas gravadas. No caso desta pesquisa eram três: 1. Verbos, 2. substantivos e 3. conjunções coordenadas.

As linhas de concordância eram enviadas diariamente a meu *e-mail* de acordo com os padrões previamente estabelecidos. O endereço de e-mail é: [glossa@tedm.ucl.ac.be](mailto:glossa@tedm.ucl.ac.be). E a mensagem enviada sempre com o arquivo em anexo está em inglês informando que a concordância solicitada esta anexada neste e-mail, e oferecem ajuda caso quaisquer informações sejam necessárias a respeito do Glossanet. Essas possíveis dúvidas seriam postadas por meio do e-mail destacado acima. E seria necessário que o usuário incluísse em sua requisição a referência do próprio e-mail e um número enviado pela equipe Glossanet.

Os arquivos vinham no formato .html. O corpus foi coletado entre 09 de maio de 2006 e 09 de julho de 2006.

Na seqüência observamos o uso da partícula “hay” expressando uma ação de obrigação: *...en concepto de pago mientras que **hay** que pagar las importaciones...*

Lendo a linha de concordância podemos estabelecer que a partícula “hay” (verbo haver- presente do indicativo – forma invariável) vem seguida da partícula “que” e um verbo no infinitivo, sempre expressando obrigação.

Outro exemplo interessante extraído das linhas de concordância se refere ao uso das conjunções “e” e “y” ambas com sentido de adição de idéias, equivalendo ao “e” do português.

*...nte, bajan los costes unitarios **y** alcanzan una mayor rentabilidad; el ...  
 ...Filipinas, Singapur, Tailandia, Vietnam **y** Brunei, concebida en sus ...  
 ...General sobre Aranceles de Aduana **y** comercio es un ...  
 ...Banco Internacional de Reconstrucción **y** Desarrollo fue fundado en 1945...  
 ... las fronteras políticas de un país) **y** exterior (realizado entre diferentes ...  
 ... favorecer la liquidez internacional **y** facilitar la forma de pago. El Banco...  
 ... las que destacan la Ronda de Tokio **y** la de Uruguay. La Conferencia de...  
 ...libre de restricciones cuantitativas **y** medidas que implican el movimiento ...  
 ...dumping sobre los productos importados) **y** no tarifarias (cuotas, licencias...  
 ... (en la actualidad casi un centenar **y** otros treinta más que aplican de facto...  
 ... y competitivo, bienestar social **y** pleno empleo para los ciudadanos de...  
 ... la adhesión de Austria, Finlandia **y** Suecia en 1995. Los principios ...*

... Cuenta con 170 países miembros y tiene como tarea principal fomentar el...  
 ... entre Argentina, Brasil, Paraguay y Uruguay que, a partir de 1995 se ...

Lendo as catorze linhas de concordância apresentadas acima, podemos identificar que depois da conjunção “y” aparecem palavras que se iniciam por quase todas as letras do alfabeto (a, b, c, d, e, f, l, m, n, o, p, s, t, u), dessa forma fica mais fácil para um aprendiz de Língua espanhola perceber que a conjunção “y” é usada muito mais do que a conjunção “e”. Isso se deve a uma regra gramatical que seria muito mais fácil de ser explicada com o auxílio do corpus e das linhas de concordância.

A regra diz que usaremos a conjunção “y” para unir palavras ou idéias, quando a palavra que segue a conjunção não se iniciar por “i” nem por “hi”. Evitando assim a junção dos sons. O uso da conjunção “e” pode ser observado na linha de concordância: ...denomina comercio a la compra, venta e intercambio de bienes y servicios con...

Outro exemplo bastante relevante para o material proposto para o Curso de Comércio Exterior é o caso do substantivo “desarrollo” que equivale a palavra “desenvolvimento” do português. Acredito que a partir da leitura das linhas de concordância, o aprendiz é capaz de inferir seu significado sem auxílio de um dicionário.

...de enero de 1993, aspira a un mayor desarrollo económico y competitivo,...  
 ... Internacional de Reconstrucción y Desarrollo fue fundado en 1945. Su ...  
 ... como de los no miembros en vías de desarrollo (PVD)". Las obligaciones ...  
 ...Naciones Unidas para el Comercio y el Desarrollo es un órgano de la ...  
 ...realizado entre diferentes países). El desarrollo del comercio exterior ...  
 ... al FMI. Su finalidad es cooperar al desarrollo de los países afiliados, ...

Observamos nas linhas de concordância acima que “desarrollo” é algo que está associado a economia, ao comércio, aos países afiliados, e que existem órgãos de reconstrução e de comércio associados ao “desarrollo”. Parece que seria mais fácil inferir seu significado e também mais significativo para o aprendiz. Já que este estaria participando da construção do significado ativamente, o que é muito mais relevante do que abrir o dicionário, verificar seu significado e depois não se lembrar mais do que se tratava.

## 6. Considerações Finais:

Por meio deste trabalho objetivou-se mostrar a importância da utilização metodológica de uma Lingüística de Corpus e do uso de linhas de concordância, baseada na aplicação do software Glossa Net, que possibilitou a elaboração de algumas considerações no que diz respeito ao uso da língua espanhola em textos jornalísticos.

Frise-se que o trabalho apresenta e analisa apenas uma pequena amostra do que pode ocorrer dentro de uma aplicação mais abrangente do corpus e das concordâncias. Este trabalho não se conclui aqui, ele tem aspectos a serem desenvolvidos de agora em diante, como por exemplo, a aplicação de atividades que se valham desta metodologia e a verificação de sua aceitabilidade.

Para finalizar, cremos que nossa pesquisa oferece uma contribuição muito importante, não só para a Lingüística Aplicada como também para a construção de metodologias de ensino mais autênticas. Estaremos sempre caminhando, trilhando novas estradas, à procura de ações pedagógicas cada vez mais informadas e comprometidas com a constituição dos cidadãos.

## 7. Referências:

Berber Sardinha, A. P. *Lingüística de Corpus*. Barueri, SP: Manole, 2004.

\_\_\_\_\_. *Lingüística de corpus: Histórico e problemática*. DELTA, 16 (2): 323-367, 2000.

Dister, Anne & Fairon, Cédric. *Extension des ressources lexicales grâce à un corpus dynamique. Lexicometrica*. Numéro thématique. Actes du colloque "L'analyse de données textuelles: De l'enquête aux corpus littéraires". Université de Louvain, 2000. Disponível on-line: <http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema7/Texte-Dister.pdf>

Fairon, Cédric & Courtois, Blandine *Les corpus dynamiques et GlossaNet*. Extension de la couverture lexicale des dictionnaires électroniques anglais. Paris: LADL, 2000.

<http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2000/pdf/52/52.pdf>

Fletcher, William H. *Concordancing the web: Promise and Problems, Tools and Techniques*. USA: United States Naval Academy, 2005.  
<http://www.kwicfinder.com/FletcherConcordancingWeb2005.pdf>

Rocha, Paulo Alexandre & Santos, Diana. *CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa*, 2000.  
<http://www.linguateca.pt/Diana/download/RochaSantosPROPOR2000.rtf>

Sinclair, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

---