A LINGUÍSTICA DE CORPUS ALIADA AO ENSINO/AQUISIÇÃO DE L2

Lívia Pretto Mottin

PUCRS

1 Introdução

A Linguística de Corpus (LdC) é uma metodologia linguística que se utiliza de textos reais. Em função do caráter autêntico dos textos que compõem suas bases de dados, a LdC surgiu como uma área da linguística que possibilita ao pesquisador o acesso, através de ferramentas computacionais, a uma vasta e rica quantidade de dados nunca antes disponível de maneira tão rápida e confiável. Uma das áreas de pesquisa linguística que se ocupa dos dados provenientes de *corpora* é o ensino/aquisição de L2. Desde o advento da LdC, informações geradas tanto por *corpora* de língua geral quanto por *corpora* de aprendizes tornaram-se fontes férteis de evidências para o desenvolvimento de material pedagógico direcionado ao ensino de línguas. Nas próximas seções, abordo alguns conceitos básicos da LdC e a forma como a área de pesquisa sobre ensino/aquisição de L2 pode valer-se das informações oriundas de *corpora*.

2 Linguística de Corpus

Granger (2002) define e LdC como uma metodologia linguística que se utiliza de textos autênticos produzidos em contextos reais de uso e que, por este motivo, é uma ferramenta poderosa no estudo e análise da língua. Essas características fazem com que a LdC tenha o potencial de mudar perspectivas sobre a língua. A autora ainda salienta que a LdC não é nem um novo ramo da linguística e nem uma nova teoria linguística. Seguindo essa mesma discussão, Rayson[1] (2002, citado por SARMENTO, 2008, p. 24) diz que "a LdC não é um ramo da linguística como a sintaxe, a semântica ou a pragmática, que concentram-se na descrição ou explicação de algum aspecto da língua em uso". Já Sarmento enfatiza que "a LdC é uma metodologia que pode ser aplicada a

[1] RAYSON, Paul. **Matrix:** A statistical method and software tool for linguistic analysis through corpus comparison. Tese de doutorado. Universidade de Lancaster, 2002.

uma grande variedade de estudos linguísticos, ou ainda ao ensino de línguas, ou seja, é uma das várias maneiras de fazer linguística" (p. 24).

Berber Sardinha (2004) acrescenta a questão do critério na compilação de *corpora* e do uso de ferramentas computacionais para suas análises em sua definição. De acordo com o autor, a Linguística de Corpus recebe a seguinte definição:

> 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. (p. 3)

Ainda segundo Berber Sardinha (2004), um *corpus* é:

> Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados de acordo com 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 descrição e análise (p.18).

A definição acima (BERBER SARDINHA, 2004) faz referência às características básicas e importantes de um *corpus*, as quais devem ser levadas em consideração tanto na compilação quanto nas pesquisas com *corpora*. As características são as seguintes: (i) representatividade; (ii) amostragem; (iii) formato eletrônico; e (iv) autenticidade.

A (i) representatividade de um *corpus* está intimamente associada ao seu tamanho (REPPEN, 2010). Na compilação de *corpora* para a produção de dicionários, por exemplo, o tamanho precisa ter proporções muito grandes a fim de incluir as mais diferentes palavras e os vários sentidos de uma mesma palavra. Desta forma, ao conduzir a investigação, o pesquisador precisa fazer a si mesmo a seguinte pergunta: "have I collected enough texts (words) to accurately represent the type of language under investigation?" (REPPEN, 2010, p.32). Reppen ainda salienta que em algumas situações, a língua sendo estudada permite que o investigador compile um *corpus* que a

represente em sua completude. Um *corpus* de músicas de uma determinada banda, por exemplo, tem a possibilidade de incluir todas as composições da banda, atingindo assim representatividade plena.

Quanto à (ii) amostragem, a função de um *corpus* é servir de amostra da língua sendo investigada. Tognini-Bonelli (2010) descreve a amostragem como sendo a característica de, através de uma amostra, representar com precisão uma variedade linguística, mostrando com exatidão as mesmas características encontradas em situações normais e reais de uso. Representatividade e amostragem estão intimamente associadas uma à outra, pois

> o corpus é uma amostra de uma população cuja dimensão não se conhece (a linguagem como um todo). Desse modo, não se pode estabelecer qual seria o tamanho ideal da amostra para que represente essa população. Uma salvaguarda é tornar a amostra a maior possível, a fim de que ela se aproxime ao máximo da população da qual deriva, sendo portanto mais representativa (BERBER SARDINHA, 2004, p. 23).

O (iii) formato eletrônico é outra das quatro características básicas e importantes de um *corpus*, tanto que "the term corpus is now almost synonymous with the term machine readable corpus" (MCENERY & WILSON, 1996, p. 17). A formatação eletrônica dos *corpora* permite que os dados sejam lidos e processados por computadores facilitando sua manipulação por parte do pesquisador. A última das características diz respeito à (iv) autenticidade. De acordo com Berber Sardinha (2004, p. 19), "o corpus deve ser composto de textos autênticos, em linguagem natural. Assim, os textos não podem ter sidos produzidos com o propósito de serem alvo de pesquisa linguística".

Se as quatro características acima citadas e descritas forem consideradas na compilação de *corpora* e no desenvolvimento de investigações linguísticas, a combinação e o uso de ferramentas computacionais com os dados de *corpora* em pesquisas linguísticas têm a possibilidade de gerar resultados quantitativos e qualitativos confiáveis, revelando fenômenos desconhecidos sobre a língua com rapidez e facilidade na manipulação dos dados. Os resultados quantitativos são estatísticos e mostrarão, por exemplo, o número de ocorrências de determinada palavra em um *corpus*. Entretanto, as pesquisas com *corpor*a podem também gerar resultados

qualitativos que permitem observar como palavras ou conjuntos de palavras são usados dentro de um contexto.

O poder de ferramentas computacionais aliado às evidências empíricas provenientes de *corpora* possibilita que o processamento dos dados seja preciso e consistente, evitando a interferência humana e dando assim maior confiabilidade à pesquisa linguística. Além disso, o caráter empírico dos estudos baseados em *corpora* permite que se chegue a resultados dos quais a intuição sozinha não daria conta (MCENERY *et al*., 2006).

As mais diversas abordagens linguísticas têm a possibilidade de utilizar a LdC em suas pesquisas. Biber *et al.* (1998) argumentam que "almost any area of linguistics can be studied from a use perspective and the corpus-based approach provides a suite of tools and methods that are particularly effective for such investigations" (p. 12).

Isso acontece porque o estudo da língua através de exemplos reais de uso proporciona acesso ao que de fato ocorre natural e autenticamente em situações de utilização da língua. Segundo McEnery *et al.* (2006), diferentemente da análise baseada na intuição do pesquisador, uma abordagem baseada no uso de *corpus* fornece evidências daquilo que os usuários acreditam ser aceitável na língua, sem qualquer julgamento de outrem. Além disso, ao se utilizar de exemplos para apoiar ou negar determinado argumento, o pesquisador está, de certa forma, monitorando a utilização que o usuário faz da língua. A intuição até pode, em algumas situações, não falhar. No entanto, pode não estar representando o uso típico, real e autêntico daquilo que aconteceria em um contexto de comunicação. Desta forma, as evidências empíricas provenientes do uso de *corpora*, fornecem ao pesquisador informações confiáveis às quais a introspecção sozinha não seria capaz de chegar.

O argumento a favor dos estudos baseados em *corpora* de Biber *et al.*(1998) é o seguinte:

> from this perspective, we can investigate how speakers and writers exploit the resources of their language. Rather than looking at what is theoretically possible in a language, we study the actual language used in naturally occurring texts (p. 1).

Como já mencionado anteriormente, um *corpus* é uma coletânea de textos autênticos armazenados e acessados através de computadores. O conteúdo dos *corpora*

é acessado através de ferramentas computacionais especializadas para tal tarefa. Alguns *corpora* estão disponíveis na internet e dispõem de seus próprios recursos de pesquisa online, por exemplo o COCA (*Corpus of Contemporary American English),* um *corpus* de língua geral representativo do inglês americano, que tem cerca de 425 milhões de palavras, foi compilado entre os anos 1990 e 2011 e é subdividido em *corpora* menores de diferentes gêneros: oral, ficcional, revistas populares, jornais e acadêmico. Já outros necessitam de programas computacionais desenvolvidos especialmente para isso. Um desses programas é o *Wordsmith Tools* (SCOTT, 2010).

As principais ferramentas disponíveis e utilizadas nas análises e pesquisas com *corpora* são: (i) lista de frequência; (ii) concordância; e (iii) lista de colocados.

As (i) listas de frequência de palavras apresentam resultados quantitativos do termo de busca e possibilitam o acesso e a identificação do vocabulário de uso comum na língua. A ferramenta apresenta uma lista das palavras do *corpus* (*types*) e o número de ocorrências de cada uma delas (*tokens*). Essa lista pode ser tanto ordenada a partir da palavra mais frequente até a mais rara, quanto organizada alfabeticamente.

Ao buscar pela palavra *thing*, o COCA, um *corpus* já mencionado anteriormente e utilizado nesta pesquisa, indica que, dentre as 425 milhões de palavras que o compõem, o termo de busca ocorre 199448 vezes. Refinando um pouco mais a pesquisa, pode-se verificar como essas ocorrências se dividem entre os cinco subcorpora de linguagem oral, ficcional, revistas populares, jornais e acadêmico, como aparece na figura 1.

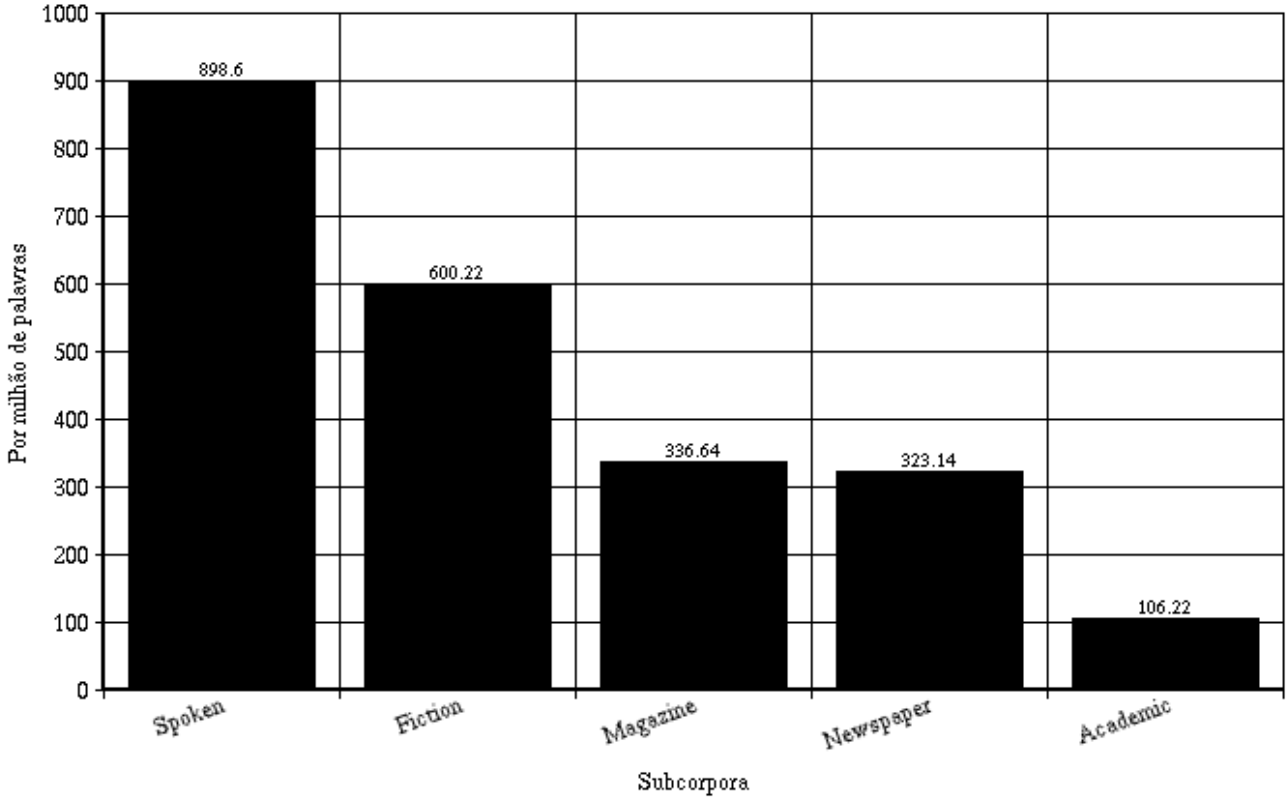

Figura 1 Frequência de *thing* nos subcorpora do COCA

A figura 1 mostra que dentre os cinco subcorpora, *thing* ocorre com mais frequência na oralidade. Além de possibilitar a busca por palavras específicas, a ferramenta também possibilita o acesso às palavras mais frequentes do *corpus*. A tabela 1 mostra as palavras com maior número de ocorrências no COCA.

Granger (2002) afirma que as evidências oferecidas por *corpora* complementam tantas outras. Entretanto, salienta que no que diz respeito à frequência, a LdC é a única fonte confiável de evidências desta natureza. Ainda no que se refere à frequência, Granger (2002, p. 4) a define como "an aspect of language of which we have very little intuitive awareness but one that plays a major part in many linguistic applications which require a knowledge not only of what is possible in language but what is likely to occur".

Com relação ao (ii) concordanciador, O'Keeffe *et al.* (2007, p. 8) argumentam que "concordancing is a core tool in corpus linguistics and it simply means using corpus software to find every occurrence of a particular word or phrase". A palavra ou

expressão de busca é chamada de palavra nódulo (*node)* e aparece no centro da tela acompanhada de cerca de oito palavras que se encontram imediatamente à sua direita e à sua esquerda (o co-texto da palavra ou expressão de busca). A tela gerada pelo programa é chamada de KWIC (Key-Word-In-Context). Cada uma das linhas apresenta um uso diferente da palavra nódulo, empregada por um falante diferente, em tempo e contextos também distintos.

A figura 2 mostra a palavra de busca *thing* ao centro, acompanhada pelo seu co-texto. Como é possível observar, o co-texto está ordenado alfabeticamente a partir da primeira palavra à esquerda da palavra nódulo.

Figura 2 Linhas de concordância de *thing* no COCA

As (iii) listas de colocados permitem a identificação das combinações de palavras com alta frequência de uso. Portanto, essa ferramenta possibilita a identificação das palavras que tendem a co-ocorrer com o termo de busca. Hunston (2002) cita o exemplo da palavra *toy* (brinquedo) que co-ocorre com frequência com *children* (crianças), diferentemente de *men* (homens) ou *women* (mulheres) que co-ocorrem junto a *toy* menos frequentemente.

Como já mencionado anteriormente, dentre os cinco subcorpora do COCA, a palavra *thing* ocorre com mais frequência na linguagem oral. Através de uma análise destas ocorrências em contexto, observa-se que na oralidade a palavra coloca-se com

outras palavras e forma expressões fixas como *that sort of thing*, *things like that* e *the thing is*, como se pode observar nas linhas de concordância da figura 3.

hard to his voice out through the stereo and **that sort of thing**. It's hard to hear sometimes. But the lines
We just don't have time in our lives for **that sort of thing**. We really enjoy the show. I love watching it.
passengers with us, women and children and all **that sort of thing**. You know? Get them on. Get them
they are rooted to the spot. And there are little **things like that**. Also they can't, they don't talk directly
more risk and higher yield. I know, when you hear **things like that** you are like, is there rope anywhere
guys doing it on the weekends for fun and **things like that** and spectators are there. It's on a weekend
a disconnect there. But, you know, **the thing is**, you have a bunch of college kids. I understand that
Good, good. But still, you know, **the thing is**, he didn't try to hurt anybody, but the truth is,
can be caught up in a destructive process. **The thing is**, though, that I don't agree with that, actually

Figura 3 Colocações de *thing* em contexto

As ferramentas citadas podem ser utilizadas na análise de diferentes tipos de *corpora*, entre os quais estão os *corpora* de aprendizes, que serão abordados na próxima seção.

3 *Corpora* de aprendizes e suas contribuições aos estudos sobre aquisição de L2

A partir de 1980, com o surgimento dos *corpora* de aprendizes, duas áreas que até então caminhavam separadamente, linguística de corpus e aquisição de L2, passaram a ver a possibilidade de terem seus campos de pesquisa trabalhando conjuntamente.

Os *corpora* de aprendizes são coleções de textos autênticos (escritos ou orais) produzidos por falantes de uma LE em uma situação de aprendizagem. Granger (2002) sugere a adoção da seguinte definição baseada em Sinclair[2]:

[2] SINCLAIR, John. **Preliminary Recommendations on Corpus Typology.** EAGLES. Disponível em: http://www.ilc.pi.it/EAGLES96/corpustyp/corpustyp.html,1996.

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance (p. 7)

Em 2009, Granger inclui em sua definição a importância de o analista estar bem preparado para realizar análises baseadas em *corpora* de aprendizes:

> Learner corpora (LC) are electronic collections of foreign or second language learner texts assembled according to explicit design criteria. The fact that they contain data from language learner makes them a very special type of corpus, requiring from the analyst a wider range of expertise than is necessary for native corpora (p. 14-15).

Os *corpora* de aprendizes oferecem uma base empírica às pesquisas sobre aquisição de L2 nunca antes disponíveis. Por este motivo, oportunizam a identificação das dificuldades dos aprendizes e têm grande potencial de proporcionar informações, descrições e percepções valiosíssimas à área de estudos sobre aquisição de segunda língua que até então, se valia de dados experimentais e introspectivos, os quais por possuírem variáveis difíceis de serem controladas, se limitavam a quantidades relativamente baixas de dados provenientes de um número também baixo de informantes, levantando, assim, questões sobre a generalização dos resultados alcançados (GRANGER, 2002). Ademais, os dados são provenientes de um número muito alto de aprendizes, tornando os *corpora* representativos daquilo que se está buscando. Podemos então argumentar em favor do uso de *corpora* de aprendizes devido à possibilidade que estes têm de superar algumas dificuldades até então enfrentadas nas pesquisas sobre aquisição de L2.

Nas palavras de Granger (2002, p. 5),

> learner corpora provide a new type of data which can inform thinking both in SLA (Second Language Acquisition) research, which tries to understand the mechanisms of foreign/second language acquisition, and FLT (Foreign Language Teaching) research, the aim of which is to improve the learning and teaching of foreign/second languages.

Esse tipo de investigação proporciona percepções valiosas ao ensino de LEs. Assim como Granger (2002), Biber *et al*. (1998) também salientam a importância dos *corpora* de aprendizes aos estudos sobre aquisição de L2:

> In the past, language acquisition was typically investigated through detailed case studies that relied on a small number of subjects. Now, as corpora of learners' language are compiled, studies can be based on a large number of learners, and general patterns across learners can be examined (p. 11-12).

Granger (2009) frisa ainda que esta é uma área de estudos que está longe de ter atingido maturidade, mas que a cada dia mais, tem se observado de uma maneira mais realista os fins aos quais os *corpora* de aprendizes podem ser aplicados, bem como suas contribuições tanto à aprendizagem quanto ao ensino de LEs.

4 A Linguística de corpus e o ensino de L2

Não há dúvidas de que, nos últimos anos, a LdC tem influenciado o ensino de línguas estrangeiras. Além de ter revolucionado a produção de dicionários, muitos dos quais, a partir do uso de *corpora*, passaram a ser baseados em *corpora* atualizados constantemente, a LdC teve um consequente reflexo no desenvolvimento de material pedagógico direcionado ao ensino de LEs. Meunier (2002, p. 124) argumenta que

> corpus research has been highly influential in initiating profound changes in reference tools. Most striking perhaps are the changes in dictionaries which, in addition to the usual lexical and grammatical information, now also provide frequency and register information in the form of language/usage notes illustrating, among other things, differences between spoken and written language.

Segundo O'Keeffe *et al.* (2007, p. 21), "in terms of what we actually teach, numerous studies have shown us that the language presented in textbooks is frequently still based on intuitions about how we use language, rather than actual evidence of use".

Entretanto, no que se refere à relação entre a LdC e o ensino de L2, é necessário que haja um equilíbrio entre dados produzidos por falantes nativos e dados produzidos por aprendizes (GRANGER, 1998). Entretanto, é sempre importante que haja um equilíbrio entre dados provenientes de nativos e de aprendizes. O material didático desenvolvido com base no equilíbrio ressaltado por Granger (1998) proporcionaria informações tanto a respeito do que é típico e recorrente na língua alvo (dados de

corpora de falantes nativos) quanto a respeito das dificuldades enfrentadas por aprendizes no processo de aprendizagem da língua alvo (dados de *corpora* de aprendizes).

Além de Granger, outros pesquisados também defendem o uso de dados de *corpora* de aprendizes no processo de aprendizagem de uma L2. Meunier (2002, citado por MCENERY *et al.*, 2006) ressalta que o equilíbrio entre os dados possibilita o acesso às possíveis lacunas existentes entre a interlíngua do aprendiz e a língua alvo. Ainda no que se refere aos estudos sobre interlíngua, Granger (1998) destaca o método contrastivo de análise dos dados. Porém, o método ao qual a autora faz referência não é o método contrastivo em seu sentido tradicional que tem como objetivo comparar/contrastar línguas diferentes. O método por ela citado tem como objetivo comparar aquilo que os falantes nativos e os falantes não nativos de uma determinada língua fazem em uma situação comparável de uso. Tal análise pode ser feita principalmente através das duas maneiras seguintes: (i) comparar língua nativa com interlíngua; e (ii) comparar interlínguas diferentes (MOTTIN, no prelo).

5 A Linguística de Corpus como ferramenta no ensino de L2

Ferramentas utilizadas no processamento de *corpora* como lista de frequência de palavras, concordanciador e lista de colocados podem ser aplicadas de diversas formas e amplamente utilizadas no desenvolvimento de material pedagógico direcionado ao ensino de línguas estrangeiras. Esta seção apresenta algumas destas aplicações.

5.1 Lista de frequência

Como já mencionado anteriormente, uma das ferramentas computacionais que permite pesquisas rápidas e confiáveis em *corpora* é a lista de frequência de palavras. Através desta ferramenta torna-se possível a obtenção das palavras de um *corpus* em ordem de frequência.

As listas de frequência são úteis no ensino pois “help us make choices about what to teach and in what order” (MCCARTEN, 2007, p. 4). Assim, assumem papel importante no ensino de vocabulário, pois permitem a priorização das palavras mais frequentes e, portanto, mais relevantes de serem incluídas no material. Entretanto, a frequência dos itens não é o único ponto que deve ser levado em consideração na ordem

de ensino de vocabulário. Outro fator que também merece atenção é o valor cultural das palavras para os aprendizes já que "some may be culturally inappropriate, not suitable for the class, or just difficult to use until students have more English" (MCCARTEN, 2007, p. 4). McCarten também destaca, além dos fatores já citados, as necessidades comunicativas dos aprendizes. Portanto, em função dos pontos acima mencionados, "frequency information, while important, is only a guide" (MCCARTEN, 2007, p. 4).

As listas de frequência também possibilitam o acesso às variações existentes entre diferentes registros. Uma mesma palavra ou expressão pode ser pesquisada em um *corpus* de língua geral e seu uso comparado aos resultados exibidos em diferentes subcorpora (subcorpora de linguagem oral, subcorpora de linguagem escrita, subcorpora de linguagem acadêmica, entre outros), por exemplo. O COCA é um exemplo *corpus* que permite a análise do uso de itens em seus cinco subcorpora, disponibilizando assim o acesso ao que ocorre na língua em diferentes contextos.

As informações geradas em buscas desta natureza são importantes, pois através delas pode-se decidir qual o contexto mais apropriado em que certa palavra deve ser apresentada aos alunos. A palavra *mean*, por exemplo, é muito mais comum na oralidade do que na escrita, como se observa na figura 4.

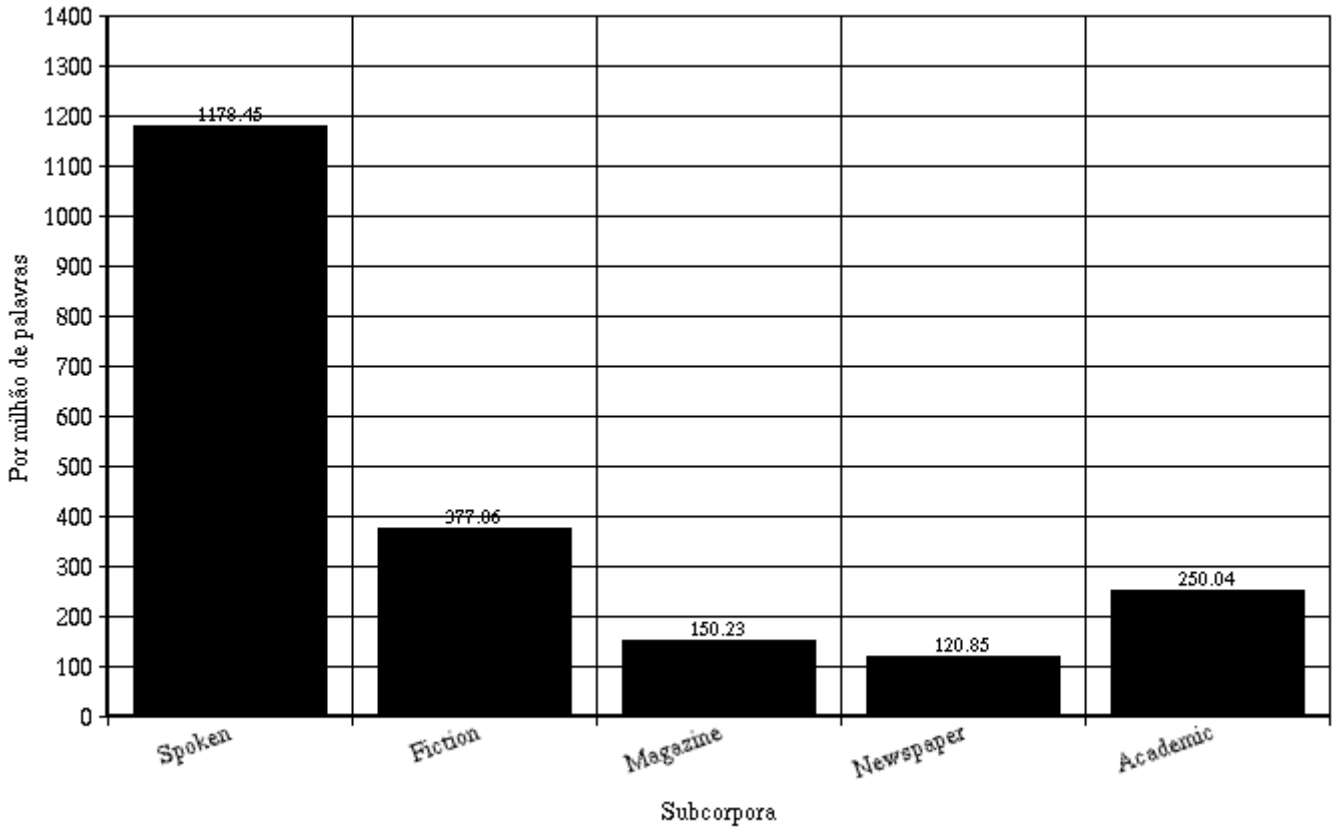

Figura 4 Frequência de *mean* nos subcorpora do COCA

5.2 Concordâncias

Ainda tendo como exemplo a palavra *mean*, a primeira busca (figura 4) apresenta resultados estatísticos, portanto, quantitativos. Se realizada de outra forma, a pesquisa gera um resultado qualitativo, mostrando as ocorrências da palavra de busca em contexto. A análise das linhas de concordância mostra que na oralidade, *mean* é uma palavra normalmente utilizada como marcador discursivo e tende a aparecer como parte da expressão *I mean* tanto para repetir ideias já mencionadas quanto para dizer mais sobre algo, sendo uma das expressões mais utilizadas na forma oral da língua inglesa.

Com base em informações como estas, podem ser desenvolvidas atividades que foquem em aspectos recorrentes e comuns na oralidade, proporcionando aos aprendizes oportunidades de aperfeiçoar habilidades comunicativas e encorajar usos que conferem fluência e naturalidade durante interações orais (MOTTIN, no prelo).

No desenvolvimento de atividades com uso de linhas de concordância, pode-se, entre tantas outras possibilidades, apresentar aos alunos atividades em que, através da observação das concordâncias de uma determinada palavra ou expressão, eles tenham que julgar se as ocorrências da palavra de busca estão sendo utilizadas em seu sentido literal ou em seu sentido metafórico, como no exemplo (figura 5).

Here are some random examples from a computer database containing lines from real language. How many examples use **hand** as a metaphor, and how many use the word **hand** in its literal sense? Use a dictionary if necessary.

1. folding the pages together by **hand**. This could take all night, so some of the
2. coconut sambol. Cigarette in **hand**, Lal would regale everyone with the exploits of his
3. Herr Oberst Sigel, on the other **hand**, rolled along in a carriage behind his men
4. I gripped my mother's **hand**, I remember, and asked what that was.
5. to have. He pressed her **hand** passionately, in the street, twisted his mouth,
6. through the 166 pages, dictionary always in **hand**. But on a Saturday afternoon long
7. has an interesting duality. On the one **hand**, Homo sapiens is just another species
8. to have the power of government in my **hand**; I am not interested in personal power
9. such prodigies are almost always taken in **hand** by one or both parents and pushed more
10. and disease, and how soap and **hand** washing could prevent its spread

Figura 5 Linhas de concordância de *hand*

5.3 Lista de colocados

A célebre frase de Firth "You shall know a word by the company it keeps" descreve bem o que uma lista de colocados pode revelar em relação a uma palavra ou expressão. Ou seja, pode-se, de certa forma, traçar o "perfil" de uma palavra através da observação das palavras que co-ocorrem junto a ela. O ensino de palavras em *chunks*[3], além de promover a fluência dos aprendizes, facilita o armazenamento do vocabulário (MOTTIN, no prelo).

Segundo O'Keeffe *et al.* (2007, p. 63),

> multiword phenomena are a fundamental feature of language use. 'Off-the-peg' vocabulary enables fluent production in real time, and would seem to be at least as significant as single-word vocabulary when it comes to investigating either the semantics or the pragmatics of the language.

Na produção de atividades para serem utilizadas em sala de aula, uma das ideias é selecionar os colocados que tendem a co-ocorrer com a palavra de busca e, em seguida, apresentar linhas de concordâncias com lacunas no lugar dos colocados, solicitando aos aprendizes que preencham as lacunas com o colocado apropriado. No exemplo abaixo, foram selecionadas 9 palavras que aparecem frequentemente uma posição à direita de *get* (figura 6). Em seguida, foram apresentadas linhas de concordância (figura 7) para serem preenchidas com os colocados mencionados anteriormente (figura 6).

	Colocado
1.	Out
2.	Into
3.	Back
4.	Some
5.	Them
6.	Away
7.	Involved

[3] A palavra *chunks* é aqui usada para designar sequências de palavras, colocações, frases prontas ou sentenças comuns em certas situações recorrentes na linguagem de falantes nativos, as quais proporcionam naturalidade ao discurso.

8.	Married
9.	Better

Figura 6 Colocados de *get*

Boy meets girl, they fall in love. They **get** _____. They have children. Nothing new here
her best friends house, along with her friend, to **get** _____ clothes for her to spend the night
He opened his eyes and yelled, " **Get** _____ of here! Leave me alone, damn you! Leave me alone!
I think the security situation will **get** _____.
" he said. " I know a place we can go to **get** _____ from them! " Preacher had gotten only a
Meanwhile, this woman has been listening from a distance and tries not to **get** _____.
few minutes before needing to pack up the kids, grab a hamburger, and **get** _____ to the hotel.
If you get hungry while you wait for your friend to **get** _____, we can have lunch sent to the office.
Panic gripped her, and she grew more solid. She had to **get** _____ the house. She ran, slipping

Figura 7 Linhas de concordância de *get* acompanhado de alguns colocados frequentes

6 Considerações finais

A LdC surgiu trazendo consigo uma nova maneira de fazer linguística. O uso de grandes bases de textos autênticos produzidos em contextos reais de uso aliado às diversas ferramentas computacionais disponíveis para a análise de *corpora* tem o potencial de revelar fatos até antes desconhecidos sobre a língua, mudando perspectivas e oferecendo novas formas de análise e pesquisa linguística.

As principais ferramentas utilizadas em pesquisas com *corpora* (lista de frequência, listas de colocados e concordância) permitem o acesso a uma riqueza de dados antes indisponível e que pode ser utilizada por pesquisadores, estudantes, professores e aprendizes de L2.

Quando aplicada ao ensino de LEs, a LdC, por proporcionar acesso ao que de fato ocorre na língua em uso, abre as mais diversas possibilidades no desenvolvimento e criação de material pedagógico. Diferentemente do material desenvolvido com propósitos puramente pedagógicos, o material didático criado com base em evidências de *corpora* reflete a língua como é utilizada por falantes nativos e ensina o que, de fato,

ocorre fora da sala de aula. Além disso, auxilia os aprendizes a soarem mais fluentes em suas interações, permitindo que eles saiam do ambiente escolar melhor preparados para enfrentar a língua alvo como é falada em contextos reais.

Segundo McCarthy (2004), a principal diferença entre o material baseado em *corpora* e os outros tipos de material é que se os aprendizes tiverem exemplos autênticos da língua alvo, não existe a necessidade de viver em contexto de imersão para experimentar a língua como ela é.

Pensando em proporcionar aos aprendizes um ensino que os torne capazes de se comunicar com eficiência na língua alvo, acredito que a LdC entra no contexto de ensino de LEs como mais uma ferramenta para ajudar no processo de aprendizagem. Entretanto, sugiro que os *corpora* não sejam utilizados por si só em contextos pedagógicos. Acredito que combinados a outras ferramentas e à inestimável mediação do professor, sejam capazes de gerar resultados ainda melhores no processo de ensino/aprendizagem de LEs.

Referências

BERBER SARDINHA, Tony. **Linguística de Corpus.** São Paulo: Editora Manole, 2004.

BIBER, Douglas.; CONRAD, Susan.; REPPEN, Randi. **Corpus Linguistics:** Investigating Language Structure and Use. New York: Cambridge University Press, 1998.

DAVIES, Mark. **The Contemporary Corpus of American English.** Disponível em: http://corpus.byu.edu/coca/

GRANGER, Sylviane. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, Sylviane. **Learner English on Computer.** Longman, 1998.

GRANGER, Sylviane. A Bird's eye view of learner corpus research. In: GRANGER, Sylviane; HUNG, Joseph; PETCH-TYSON, Stephanie (Editors). **Computer learner corpora, second language acquisition, and foreign language teaching.** Amsterdam: John Benjamins, 2002. p. 3-33.

GRANGER, Sylviane. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In: AIJMER, K. (Ed.). **Corpora and Language Teaching.** Amsterdam: John Benjamins, 2009. p. 13-32.

HUSTON, Susan. **Corpora in Applied Linguistics.** London: Cambridge University Press, 2002.

MCCARTEN ,Jeanne. **Teaching Vocabulary:** Lessons from the Corpus, Lessons for the Classroom. Cambridge: Cambridge University Press, 2007.

MCCARTHY, Michael. **Touchstone:** From Corpus to Course Book. Cambridge: Cambridge University Press, 2004.

MCENERY, Tony; WILSON, Andrew. **Corpus Linguistics:** An introduction. Edinburgh: Edinburgh University Press, 1996.

MCENERY, Tony; XIAO, Richard; TONO, Yukio**. Corpus-Based Language Studies –** an advanced resource book. Oxon: Routledge Taylor & Francis Group, 2006.

MEUNIER, Fanny. The pedagogical value of native and learner corpora in EFL grammar teaching. In: GRANGER, Sylviane; HUNG, Joseph; PETCH-TYSON, Stephanie (Editors). **Computer learner corpora, second language acquisition, and foreign language teaching.** Amsterdam: John Benjamins, 2002. p. 119-141.

MOTTIN, Lívia P. **As contribuições da Linguística de Corpus ao desenvolvimento de material pedagógico direcionado ao ensino de inglês como L2.** E-book da disciplina Tópicos em Aquisição de L2, ministrada pela Profa. Dr. Lilian Cristine Scherer, na Pontifícia Universidade Católica do Rio Grande do Sul. No prelo.

O'KEEFFE, Anne; McCARTHY, Michael; CARTER, Ronald. **From Corpus to Classroom:** Language Use and Language Teaching. Cambridge: Cambridge University Press, 2007.

REPPEN, Randi. Building a corpus: What are the key considerations? In: O'KEEFFE, Anne; McCARTHY, Michael. **The Routledge Handbook of Corpus Linguistics.** New York: Routledge, 2010. p. 31-37.

SARMENTO, Simone. **O uso dos verbos modais em manuais de aviação em inglês:** Um estudo baseado em corpus. Tese se doutorado. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2008.

SCOTT, M. **WordSmith Tools.** 5.0 Version. Oxford: Oxford University Press, 2010.

TOGNINI-BONELLI, Elena. The evolution of Corpus Linguistics. In: O'KEEFFE, Anne; McCARTHY, Michael. **The Routledge Handbook of Corpus Linguistics.** New York: Routledge, 2010. p. 14-27.