

## Utilizando corpora de aprendizes para a investigação de aspectos discursivos, metodologias de ensino e design de materiais pedagógicos

(Using learner corpora to investigate discursal aspects,  
teaching methodologies, and pedagogical materials)

Leonardo RECSKI  
Universidade Federal de Santa Catarina

*ABSTRACT: This paper reviews some of the corpus-based studies which have been carried out into learner corpora and their applications to the compilation and design of pedagogical materials for learners of English as a foreign language (EFL). The paper then goes on to describe how the results obtained from research on learner corpora in three specific areas, namely, discursal characteristics, collocational patterns, and adequate interpersonal tone may be adjacently employed with specialized corpora to create 'awareness-raising' EFL pedagogical materials. The paper also puts forward that the area under scrutiny has important repercussions for EFL teaching methodology. Although the paper supports the use of concordancing and learner corpora in EFL teaching some caveats in implementing this approach are also given.*

*RESUMO: Este trabalho faz uma resenha de estudos sobre lingüística de corpus que foram conduzidos usando*

## UTILIZANDO CORPORA DE APRENDIZES

*corpora de aprendizes e suas aplicações para a compilação e elaboração de materiais didáticos para aprendizes de inglês como língua estrangeira. O texto descreve como os resultados obtidos de pesquisa sobre corpora de aprendizes em três áreas específicas, ou seja, características discursivas, padrões de co-ocorrência (colocações) e o tom interpessoal adequado, podem ser empregados com corpora especializados para criar materiais didáticos. Os impactos da área para a metodologia de ensino do inglês como língua estrangeira também são discutidos. Embora se defenda aqui o uso de corpora e programas de concordancing, no ensino da língua estrangeira, algumas restrições também são apresentadas.*

*KEY-WORDS: learner corpora, pedagogical materials, teaching methodology, concordancing.*

*PALAVRAS-CHAVE: corpora de aprendizes, material didático, metodologia de ensino, concordancing.*

## INTRODUÇÃO

Nas últimas duas décadas, a área de lingüística de corpus vem sendo ampliada em pelo menos um aspecto: atualmente ela envolve uma quantidade maior, tanto de corpora de grande, como de pequena escala. Corpora de grande escala (mega-corpora) não são mais um sonho de alguns idealizadores do passado (vide Kennedy (1998) para uma ótima revisão de tais corpora). Por exemplo, o *Cobuild Corpus* vem crescendo continuamente e atualmente contém cerca de 418 milhões de palavras. Novos textos são continuamente adicionados a este

banco de textos gigantesco utilizado, entre outras coisas, na compilação de dicionários e gramáticas (Willis e Wright, 1995; Sinclair, 1995). Outro exemplo da dimensão de certos corpora contemporâneos é o *Cambridge International Corpus* (CIC), que vem sendo compilado desde o início da década de 90 e, que atualmente, contém mais de 600 milhões de palavras envolvendo diferentes gêneros discursivos. Os recursos do CIC permitem que lexicógrafos e idealizadores de materiais pedagógicos da *Cambridge University Press* tenham acesso a jornais, revistas, romances, livros, transcrições de fala, websites, programas de rádio e TV e muitos outros gêneros discursivos para a compilação de seus materiais pedagógicos. Existe também o *British National Corpus* (BNC), com seus 100 milhões de palavras (Aston e Burnard, 1998), compilado a partir de vários gêneros escritos (80%) e falados (20%), com o propósito de ser representativo do inglês britânico contemporâneo.

Outro mega-projeto iniciado em 2001 foi o *Global English Monitor Corpus*<sup>1</sup>. O *Global English Monitor Corpus* está sendo compilado com base nos maiores jornais mundiais de língua inglesa. O objetivo deste projeto é armazenar bilhões de palavras que possibilitarão investigar, por exemplo, se o inglês produzido por britânicos, americanos, sul-africanos, australianos, indianos etc, vem se modificando de maneira semelhante ou não, através da documentação do uso da linguagem e de variações semânticas refletidas nesses jornais. O projeto procura beneficiar não apenas lexicógrafos, semanticistas e analistas do discurso interessados em estudos diacrônicos, mas também sociólogos e cientistas políticos ao redor do mundo.

---

<sup>1</sup> Para maiores informações acesse o site [www.corpus.bham.ac.uk/global/htm](http://www.corpus.bham.ac.uk/global/htm).

## UTILIZANDO CORPORA DE APRENDIZES

Mesmo com esta explosão das dimensões de alguns corpora, sob um outro prisma, a área da lingüística de corpus vem se expandindo e incluindo corpora bem menores, baseados em gêneros bem mais específicos. Aston (1997) salienta que estes tipos de corpora geralmente possuem entre 20 e 200 mil palavras e são bem mais específicos do que corpora maiores em relação aos seus tópicos e gêneros discursivos. Grande parte desses corpora incluem várias áreas da ciência e tecnologia e normalmente consistem de amostras de discurso acadêmico escrito ao invés do falado<sup>2</sup>.

Materiais didáticos desenvolvidos a partir de concordâncias muitas vezes incluem o que certos profissionais da área de ensino consideram como características preponderantes da escrita acadêmica; áreas geralmente consideradas problemáticas na escrita de aprendizes; ou áreas que os próprios aprendizes identificam como sendo problemáticas (Thurstun e Candlin, 1998; Pickard, 1994; Stevens, 1991).

O uso restrito de corpora de aprendizes para o desenvolvimento de materiais pedagógicos pode estar relacionado a pelo menos dois fatos: (i) não são disponibilizados publicamente, e (ii) os corpora comercialmente disponíveis custam caro. Desta forma, professores de EFL precisam compilar seus próprios corpora, seja escaneando a produção de seus aprendizes, seja pedindo que eles armazenem esta produção em disquetes

---

<sup>2</sup> Alguns exemplos de corpora compilados para fins específicos envolvem: o corpus de inglês acadêmico escrito de Ken Hyland (1,3 milhões de palavras divididas em oito disciplinas); o corpus de Tim Johns de *Plant Biology* (100 mil palavras); o corpus de escrita britânica negra de Mark Sebba; o TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL) compilado pelo Educational Testing Service, USA (2,8 milhões de palavras) e o Corpus for English in Science and Technology (JDEST) de Jiao Tong (1 milhão de palavras extraídas de áreas como física, engenharia e tecnologia).

ou CDs. Apesar do difícil acesso a corpora de aprendizes, este artigo sugere que materiais pedagógicos poderiam ser enriquecidos, se as dificuldades “reais” destes aprendizes fossem capturadas, armazenadas em bancos textuais e posteriormente escrutinadas de forma minuciosa.

Este artigo está dividido em duas partes. Em primeiro lugar, ele revisa alguns dos estudos que vem sendo realizados com base em corpora de aprendizes de pequena escala, para então analisar algumas das aplicações pedagógicas que estão sendo propostas. A segunda parte do artigo ilustra, com alguns exemplos, como os resultados obtidos a partir de um corpus de aprendizes, i.e. sub-corpora do *International Corpus of Learner English (ICLE)*<sup>3</sup>, podem ser empregados paralelamente a corpora mais especializados na preparação de materiais pedagógicos.

## A INVESTIGAÇÃO LINGÜÍSTICA EM CORPORA DE APRENDIZES

Atualmente, vários estudos vem sendo realizados em partes de um grande corpus de aprendizes: o ICLE. Com o propósito de identificar quais estruturas gramaticais, lexicais, ou itens discursivos são empregados de forma restrita ou excessiva na escrita acadêmica de aprendizes de inglês como língua estrangeira, os resultados obtidos a partir desses sub-corpora são normalmente contrastados a um corpus paralelo

---

<sup>3</sup> O ICLE (Granger, 1993) é um projeto multinacional que busca identificar as características compartilhadas por uma grande variedade de interlínguas. Neste projeto, cada subcorpus nacional contribui com cerca de 200 mil palavras (cerca de 400 composições argumentativas).

## UTILIZANDO CORPORA DE APRENDIZES

de falantes nativos, ou às vezes a corpora referenciais de maior escala.

Vários estudos contrastivos envolvendo o componente francês do ICLE e o *Louvain Corpus of Native English Essays* (LOCNESS) como base comparativa, foram realizados pela fundadora e coordenadora do projeto ICLE Sylviane Granger (1998a) em colaboração com outros pesquisadores. O sub-corpus francês contém 280 mil palavras e consiste de composições argumentativas escritas por alunos universitários de inglês como língua estrangeira. O LOCNESS possui 230 mil palavras e é composto por textos similares aos do componente francês escritos por alunos universitários britânicos e americanos. Em um estudo contrastivo, Granger e Rayson (1998) examinaram padrões estatisticamente significantes de uso excessivo ou restrito de várias classes gramaticais (artigos, verbos, substantivos, advérbios, preposições e pronomes) no corpus dos aprendizes e descobriram não apenas um excesso estatisticamente significativo de pronomes pessoais (1ª e 2ª pessoas), mas também um uso significativamente restrito de preposições; categoria geralmente associada a nominalizações em escrita acadêmica informativa (cf. Biber et al., 1998). Entretanto, Granger e Rayson comentam que dentro da categoria preposições existem diferenças consideráveis. Por exemplo, preposições informais tais como *till* e *in spite of* são empregadas excessivamente, ao passo que preposições mais formais, como, por exemplo, *until* e *despite*, são pouco empregadas. Assim, eles concluem que a produção desses aprendizes apresenta muitas características estilísticas de um registro oral e poucas características típicas de escrita acadêmica. Estudos semelhantes, envolvendo o uso de conjunções, são descritos em Granger e Tyson (1996). Altenberg (1997) descreve vários estudos realizados com o sub-corpus sueco envolvendo áreas como

vocabulário, complexidade de grupos nominais e envolvimento pessoal.

Lorenz (1998) analisou o sub-corpus alemão do ICLE. Ele investigou o grau de intensidade vinculado aos argumentos expressos na escrita de aprendizes universitários alemães, e contrastou-os aos padrões encontrados no LOCNESS. Os resultados indicam, afirma Lorenz, que os aprendizes alemães não apenas usam muito mais advérbios de intensidade do que os aprendizes britânicos e americanos, mas também tendem a situá-los em posição temática nas orações. Os aprendizes nativos, por outro lado, tendem a intensificar adjetivos localizados dentro de grupos nominais. Recski (2004) investigou a expressão de ‘assertividade’ na escrita de aprendizes brasileiros. Ele analisou de que maneira aprendizes brasileiros expressam certeza e/ou incerteza investigando o sub-corpus brasileiro do ICLE (Br-ICLE – <http://lael.pucsp.br/corpora/bricle/index.htm>), e contrastou seus dados àqueles obtidos através da investigação do LOCNESS. Seus resultados indicam que os aprendizes brasileiros parecem ter dificuldade ao dosar a força de seus argumentos, demonstrando um grau acentuado de assertividade e um tom mais autoritário do que aquele encontrado na escrita dos alunos britânicos e americanos.

Embora ultimamente uma quantidade significativa de pesquisa venha sendo desenvolvida com base em corpora de aprendizes, poucos resultados tem sido diretamente utilizados pedagogicamente e tendem a permanecer apenas em nível de implicações. O artigo sugere que resultados de estudos comparativos entre corpora de aprendizes e corpora de falantes nativos poderiam ser utilizados para o desenvolvimento de materiais pedagógicos. Um bom exemplo disto é a compilação de dicionários para aprendizes de inglês como língua estrangeira. Por exemplo, muitas das entradas do *Longman Dictionary of*

## UTILIZANDO CORPORA DE APRENDIZES

*Common Errors* (Turton e Heaton, 1996) são baseadas na comparação entre o BNC e o *Longmans' Learner Corpus*<sup>4</sup>. A próxima seção revisa algumas das aplicabilidades de corpora de aprendizes na criação de materiais pedagógicos.

## EXPLORANDO CORPORA DE APRENDIZES

Tribble (1990, 1991) foi um dos pioneiros no uso de pequenos corpora de aprendizes. Tribble investigou o uso de verbos relacionados à fala em um corpus de aprendizes de cerca de 54 mil palavras, a maioria extraída do *Longman Corpus of Learner English*. Tribble notou um emprego excessivo e inapropriado do verbo *said* por parte dos aprendizes se comparado ao *English Historical Review Corpus*. O autor sugere que os exemplos obtidos a partir das linhas de concordância derivadas do corpus de aprendizes podem ser explorados por eles mesmos em sala de aula, de maneira que sejam utilizados para auxiliá-los na reestruturação de sentenças, o que, segundo Tribble, os incentivaria a empregar uma variedade maior de vocabulário.

Outro estudo baseado em corpora de aprendizes contendo algumas sugestões para práticas pedagógicas é o de Granger e Tribble (1998). Eles compararam o sub-corpus francês do ICLE com um sub-corpus de 1 milhão de palavras extraído do

---

<sup>4</sup> O *Longmans' Learner Corpus* (10 milhões de palavras) é compilado a partir de exemplos de escrita formal e informal de aprendizes de EFL de oito níveis de proficiência representando 160 nacionalidades.

BNC utilizando a ferramenta *Keywords*<sup>5</sup> do programa *WordSmith* (Scott, 1996). Os resultados obtidos por Granger e Tribble (1998) revelam que estes aprendizes empregam exageradamente adjetivos *multiuso* como *important*, que são privilegiados em detrimento de adjetivos com um maior grau de especificidade (e.g. *crucial*, *fundamental*).

A utilização de resultados obtidos através do estudo de corpora de aprendizes na forma de atividades baseadas em *Computer Assisted Language Learning* (CALL) propostos por Milton (1999) é provavelmente uma das mais detalhadas existentes hoje em dia. Entretanto, estas atividades demandam um grau de conhecimento muito elevado no uso de certas ferramentas computacionais e também uma grande habilidade para serem implementadas, o que, na maioria dos casos, não corresponde a realidade do típico professor de inglês como língua estrangeira. Além disso, corpora de aprendizes não são, pelo menos por enquanto, disponibilizados para professores de inglês pelo simples fato de não serem amplamente comercializados. Mesmo que fossem disponibilizados, eles seriam restritos a um contexto de ensino específico de outro lugar. Sugiro que uma boa alternativa para o professor de língua estrangeira é compilar seus próprios corpora de pequena escala; corpora que reflitam melhor o *seu* contexto de ensino.

O artigo procura demonstrar como os resultados derivados de estudos baseados em pequenos corpora compilados, podem, sem muita dificuldade, serem utilizados em aulas de inglês a nível universitário. Paralelamente, o artigo discute

---

<sup>5</sup> A ferramenta *Keywords* é utilizada para localizar e identificar palavras-chave em um dado texto. Para fazer isto, o programa compara as palavras de um corpus com as palavras de um corpus referencial (normalmente várias vezes maior do que o corpus sob investigação). Qualquer palavra considerada marcante em termos de frequência é considerada “chave”.

## UTILIZANDO CORPORA DE APRENDIZES

possíveis problemas associados ao uso de corpora em sala de aula, sugerindo alternativas para lidar com certos tipos de problemas freqüentemente encontrados na produção de aprendizes.

A próxima sessão discute como linhas de concordância e corpora de aprendizes de pequena escala podem ser utilizados paralelamente para ressaltar aspectos relacionados a três áreas de competência lingüística: características discursivas, padrões de co-ocorrência (colocações) e o tom interpessoal adequado.

## UTILIZANDO O ICLE: ALGUMAS POSSÍVEIS APLICAÇÕES

### *Características discursivas*

Uma década atrás, grande parte dos estudos baseados em corpora estavam voltados para a identificação de padrões léxico-gramaticais de textos desconsiderando-se o contexto ao seu redor (Flowerdew, 1998a); mas, atualmente, vários estudos contextualizados vem sendo realizados no intuito de revelar certas características discursivas de textos.

Em um estudo de itens lingüísticos empregados para a expressar causa e efeito, Flowerdew (1998b) analisou o uso de conjunções e substantivos empregados com a função de organizar o discurso em um corpus de aprendizes e em um corpus especializado de cerca de 40 mil palavras cada. Uma importante diferença encontrada por Flowerdew (1998b), entre a escrita dos aprendizes e o corpus especializado foi o uso excessivo de conjunções empregadas pelos aprendizes para estabelecer o que o autor chama de “local coherence” (p.343) Por exemplo, Flowerdew comenta que encontrou 32 ocorrências de *therefore* no corpus dos aprendizes, mas apenas 13 no

## LEONARDO RECKSI

corpus de escrita especializada. Além disso, ele comenta que no corpus de escrita especializada *therefore* ocorre no início das sentenças em apenas uma linha de concordância, ao passo que no corpus dos aprendizes, com exceção de uma ocorrência, todas as outras ocorrências de *therefore* ocorreram no início das sentenças.

A fim de verificar se os aprendizes brasileiros, assim como os aprendizes investigados por Flower (1998b), empregam de maneira excessiva conjunções utilizadas para expressar causa e efeito, realizei uma e os resultados obtidos foram contrastados à escrita de alunos britânicos e americanos (LOCNESS) em dois corpora de aproximadamente 33 mil palavras. Em relação a frequência, os resultados indicam que os aprendizes brasileiros empregam *therefore* de forma restrita ( $\chi^2 = 24,27$ ;  $df = 1$ ;  $p > 0,005$ ) se comparados aos aprendizes americanos e britânicos. Porém, das 22 ocorrências de *therefore* encontradas no corpus brasileiro, 17 (77%) ocorrem no início das sentenças, ao passo que no LOCNESS, onde foram encontradas 69 ocorrências de *therefore*, apenas 4 (6%) ocorrem em posição temática. Além disso, há uma ausência de conjunções tais como *thus* e *then* sendo empregadas para sinalizar resultados no corpus de aprendizes brasileiros. Estas conjunções são encontradas no LOCNESS, sendo normalmente empregadas para resumir o parágrafo que as antecede. Nas linhas de concordância abaixo, extraídas do LOCNESS, a maioria dos exemplos de *then* é empregado para servir a este propósito.

d all its attributes. Guilt and remorse are,	<b>then,</b>	very powerful and important themes in “L
illard says of Camus. It can be concluded	<b>then</b>	that not only Clamance, but also Camus fe

## UTILIZANDO CORPORA DE APRENDIZES

ding- it is simply free. If a man feels free,	<b>then</b>	he would not look to the past. He would re
er humans. If you do not believe in God,	<b>then</b>	it is this judgment of each other which is o
By his portrayal of guilt and remorse,	<b>then,</b>	Sartre is condemning such manifestations
use to the party. Events in the play show,	<b>then,</b>	that Hugo does deserve to be called an ana

Por outro lado, os aprendizes brasileiros geralmente utilizam *then* para sinalizar causa-efeito a nível local, como é evidenciado nas linhas de concordância abaixo. Estes dois usos distintos de *then* correspondem ao que Halliday e Hasan (1976) chamam de conjunções ‘internas’ e ‘externas’, onde interna significa que a conjunção está relacionada à própria interação e externa está “relacionada aos fenômenos que constituem o conteúdo que está sendo descrito” (1976, p.321; minha tradução).

ction for its members. Those members will	<b>then,</b>	have the ability to actively participate i
ind followers of the existing leaders and are	<b>then,</b>	put apart of any leadership position. Th
and emotion; if he/she likes skate boarding	<b>then</b>	he/she’ll listen to rap music. It’s always
other, teacher, etc. If the individual is gifted	<b>then</b>	the work is easier to perform. The patie
keep them arrested in waste public MONEY.	<b>Then</b>	the good solution could be submit them
asting what they need or do not need to do.	<b>Then</b>	if you do not think about what is happe
y would not agree with this proposition and	<b>then,</b>	more war for peace. That’s why I totally

## LEONARDO RECSKI

É interessante ressaltar que certos substantivos utilizados na organização do discurso, tais como *result*, *cause* e *reason*, apresentaram diferentes padrões gramaticais nos dois corpora. No LOCNESS eles aparecem no rema das orações, e.g. ... *but doesn't see that this is the result of a fundamental choice made by the individual...*. Por outro lado, os aprendizes brasileiros tendem a pré-modificar estes substantivos e posicioná-los em posição temática, como as linhas de concordância abaixo exemplificam:

you can construct and believe. Another	<b>reason</b>	for attending an university is that it is re
anyone considered "inferior". One of the	<b>reasons</b>	for that is because Sao Paulo is a metrop
w technique of youthfulness. But the best	<b>reason</b>	considered by the government is that the
to all kinds of cruelty. Maybe one of the	<b>reasons</b>	for that is the way that teenagers are mor
iends.Money keeps people together. The	<b>cause</b>	of so much poverty in the world is due t

Para que possamos salientar estes traços discursivos, é possível sugerir aos aprendizes que explorem um contexto mais amplo do que aquele oferecido nas linhas de concordância (a maioria dos software de concordância possui tal opção). Por exemplo, os aprendizes podem utilizar uma quantidade maior de contexto para examinar trechos anteriores do texto e determinar a real função que está sendo exercida por uma determinada conjunção, e.g. se ela está sendo empregada para descrever causa-efeito em nível local ou global. Neste caso, nem sempre é apropriado examinar os resultados obtidos através das linhas de concordância, seja na tela do computador, seja de forma impressa. Isto porque vários exemplos muitas vezes só podem

## UTILIZANDO CORPORA DE APRENDIZES

ser analisados com base no(s) parágrafo(s) que os antecedem. Também é possível sugerir que eles comparem o uso destas conjunções baseando-se em dois ou mais corpora, para que assim, tenham a possibilidade de analisar se estão empregando estas conjunções de forma semelhante ou diferente a de outros grupos de aprendizes.

Entretanto, é necessário que algumas palavras de cautela sejam ditas aqui. É possível que haja uma aversão ao uso de tais métodos não apenas por parte dos alunos, mas também por parte de muitos profissionais da área de ensino. É possível também, que alunos provenientes de cursos tais como arquitetura, medicina, economia, administração, engenharias etc., tenham reações variadas com relação ao uso de ferramentas computacionais e aos tópicos dos textos que são utilizados em sala de aula para o aprendizado de inglês como língua estrangeira / inglês para propósitos específicos. Portanto, a título de sugestão, acredito que uma forma de motivar os aprendizes a explorar as possibilidades oferecidas pelo uso de corpora e linhas de concordância em sala de aula é oportunizar a eles a escolha de textos que gostariam de trabalhar. Finalmente, é possível que certos alunos não aprovem a idéia de utilizarem corpora de aprendizes em seu aprendizado, pois estes corpora, provavelmente, contém erros e podem não refletir o padrão de uma 'boa' escrita acadêmica.

### *Padrões de co-ocorrência*

Inúmeros estudos baseados em corpora vêm enfocando a importância do léxico para uma escrita bem sucedida (Pickard, 1994; Flowerdew, 2000; Granger, 1998a) Com base nos resultados encontrados em alguns destes estudos, uma análise semelhante foi realizada aqui para examinar o uso de certas palavras

## LEONARDO RECSKI

em seis sub-corpora do ICLE: os sub-corpora brasileiro, francês, espanhol, tcheco, alemão e finlandês (cada um com cerca de 33 mil palavras). Algumas das palavras investigadas através do uso de linhas de concordância nestes corpora foram: *research, survey, findings, project, analysis, investigation*, etc. A análise revelou que os aprendizes conhecem estes itens lexicais, mas muitos deles não estão familiarizados com o contexto léxico-gramatical condizente a eles. Alguns erros de colocação, tais como os demonstrados nos exemplos (1) e (2), foram encontrados:

- (1) In a research made in the United States, they found out that adult people are also attracted by violence. (sub-corpus francês)
- (2) They have performed a survey to investigate if these symptoms were related to the working conditions. (sub-corpus finlandês)

Erros como os exemplificados acima foram extensivamente analisados por Howarth (2001) em um estudo contrastivo envolvendo padrões combinatórios produzidos por aprendizes de inglês como língua estrangeira e falantes nativos de inglês. Howarth estabeleceu padrões combinatórios distribuídos em três categorias: combinações livres, combinações restritas e expressões fixas. Ele observou que a categoria combinações restritas – i.e. (a) tanto o verbo quanto o substantivo podem ser substituídos (e.g. *introduce/table/bring forward a bill/an amendment*); (b) a escolha do substantivo é restrita, mas algumas substituições verbais são permitidas (e.g. *pay/take heed*); (c) há uma restrição total na escolha verbal, mas algumas substituições dos substantivos são possíveis (e.g. *give the appearance/impression*) – foi a categoria mais saliente encon-

## UTILIZANDO CORPORA DE APRENDIZES

trada na escrita dos falantes nativos. Naturalmente, ele constatou que esta categoria é a mais problemática para os aprendizes, comentando que estes tendem a cometer erros envolvendo substituições lexicais como os erros demonstrados nos exemplos (1) e (2).

Para que possamos auxiliar os aprendizes a evitarem tais erros combinatórios é possível sugerir que eles recebam exercícios no formato de listas de verbos e substantivos potencialmente problemáticos neste aspecto para que assim possam tentar combiná-los. Após receberem exercícios, através dos quais eles têm a oportunidade de testar suas hipóteses, eles podem verificar suas respostas através do uso de software de concordância para checar tais combinações em um corpus contendo textos que envolvam um gênero semelhante. Eles podem também, examinar as estruturas sintáticas prototípicas onde estas combinações ocorrem; por exemplo, verificar se a combinação *carry out a research* ocorre mais frequentemente na voz passiva ou na ativa.

Entretanto, uma das desvantagens de utilizar exercícios baseados em linhas de concordância é que às vezes os aprendizes tendem a ficarem entediados ao serem expostos a muitos exercícios desta natureza. Uma possível solução para esse problema é intercalar exercícios baseados em linhas de concordância com exercícios escritos ou discussões posteriores.

### *Negociando o tom interpessoal adequado*

Utilizando o sub-corpus brasileiro, uma investigação do tom interpessoal atrelado a proposições e argumentos expressos nas composições desses aprendizes foi realizada. O tom interpessoal está relacionado à atitude destes aprendizes em relação aos seus argumentos. Essa atitude é tipicamente exter-

## LEONARDO RECSKI

nalizada através de verbos modais (e.g. *may, should, will*), adjuntos modais (e.g. *of course, obviously, possibly*) e ítems linguísticos utilizados para intensificar (e.g. *really, completely*) ou reduzir (e.g. *fairly, a little bit*) a força de certos argumentos. Tonalidades interpessoais inadequadas muitas vezes refletem o fato do autor estar interpretando de forma incorreta tanto o contexto situacional quanto o contexto cultural onde o texto está sendo produzido. Por exemplo, em algumas composições, certos aprendizes freqüentemente empregam um tom exageradamente pessoal, como se estivessem em um bate-papo informal com amigos, quando, de fato, estão escrevendo para uma audiência acadêmica bem mais formal e criteriosa. Os exemplos (3) e (4) demonstram a informalidade com que certos aprendizes expressam seus pontos de vista:

- (3) In last week's publication of *Veja*, there was an article about girls that are between nine and twelve years old and behave like women. Instead of playing with dolls or balls they prefer to go to the hairdresser and wear a kind of clothes that make them look much older than they really are. In my opinion it happens because their parents do not care about the way they dress, they even motivate them buying women's clothing and shoes and saying they look nice wearing that sort of clothes. Another way to motivate them to behave like adults is buying them make-up and also taking them to the beauty parlor. There are some girls who prefer to spend their birthday in a hairdresser's than having a party with all their friends. I think this is a pity!
- (4) For George Orwell, the pure socialism can not exist and I totally agree with him. However, in my opinion, I would go a little bit farther. I would say that the pure socialism that professes the equality of human being isn't possible to become true.

## UTILIZANDO CORPORA DE APRENDIZES

Outros exemplos deste problemático tom interpessoal ocorrem em certos trechos das composições onde uma série de recomendações são sugeridas (exemplos (5) e (6)) Tais exemplos parecem ser socio-culturalmente inapropriados pelo fato destes aprendizes estarem se dirigindo, por exemplo, a autoridades governamentais e organizacionais, para os quais um tom menos autoritário provavelmente seria preferível:

- (5) Why there is so many beggar, so many kids asking for money at the lights, so many thieves. Not because they like to be this way (well, some of them really do), but because there is no space for everyone in the world. Space in terms of everything (hospitals, schools, jobs, food). Violence has been increasing in the past years because of that. Too many people, who can't have a decent life, seem to be blaming the ones who have, so they kill, steal, and attack people on the streets until they bleed. That's a reality not only in big cities but in small cities ass well. So, I completely think the government should limit the number of children a family can have, specially if the family can't support the children.
- (6) Television in Brazil became a synonymous of sex in the last decade. It's funny isn't it? This is exactly where the money is. Remember, showing the people what they want to see, even if it's not the right thing to do, is a good way to fill up your pocket. Showing this kind of program, like Sex scenes or insinuations, can change children mind's, giving them a wrong idea, or showing them a world that's not real. It's time for a break on this exploitation! The competent organizations should do more about that, not only keep saying that it's wrong, but act quickly.

Uma possível explicação para este exagerado tom interpessoal reside no fato de que muitos aprendizes pensam que,

desta forma, suas recomendações serão mais persuasivas. Por outro lado, muitos aprendizes acreditam que se empregarem um tom mais atenuado para expressarem seus pontos de vista, estariam, desta forma, demonstrando que suas evidências não são suficientemente embasadas para causar o impacto desejado.

Estes dois casos demonstram o quão importantes são o contexto situacional e cultural na determinação de um tom interpessoal adequado. Muitas vezes é difícil dizer, apenas examinando o texto, se a ‘intensidade pragmática’ de uma expressão é adequada ou não sem que se tenha conhecimento de um contexto cultural mais amplo, como por exemplo, a relação entre o autor e o leitor e as práticas sociais da comunidade discursiva onde o texto foi criado. Aston (1995, p.260) sugere que dados extraídos de linhas de concordância demonstram “a descontextualização de ocorrências individualizadas extraídas de seu contexto comunicativo” (minha tradução). Pelo fato do corpus estar separado do contexto situacional que o originou, Widdowson (1998, p.712) argumenta que ele pode ser visto apenas como o produto textual de um discurso pelo fato de “não poder nos informar a respeito dos processos discursivos através dos quais significados pragmáticos são apropriadamente atingidos” (minha tradução).

Estas observações têm importantes desdobramentos para aplicações pedagógicas embasadas em corpora. Widdowson (1998) argumenta que professores devem “autenticar” a língua como pertencendo ao discurso e devem relacioná-la aos seus próprios contextos culturais. Ele sugere que professores devem, de fato, “preocuparem-se com significados pragmáticos, mas isso só pode ser atingido se situarem a língua e criarem as condições textuais para torná-la uma realidade para certos

## UTILIZANDO CORPORA DE APRENDIZES

grupos de aprendizes, para que assim eles possam autenticá-la” (p.715; minha tradução).

Portanto, em algumas circunstâncias, existe um perigo inerente na transferência direta de resultados obtidos através de estudos baseados em corpora para pedagogia – o que é pragmaticamente apropriado em um corpus talvez não seja apropriado nos contextos situacionais e culturais destes aprendizes, não sendo, desta maneira, diretamente transmissível. Por essa razão, concordâncias devem ser utilizadas de forma judiciousa. Em certas circunstâncias, os aprendizes podem ser alertados com relação às suas ‘deficiências pragmáticas’, não através do uso de concordâncias, que podem oferecer informações pelo menos ‘duvidosas’, mas sim de uma forma mais indireta, através de discussões embasadas em questões culturais bem definidas, com as quais os aprendizes estejam familiarizados e que possam estimulá-los a verificar o quão adequados os seus tons interpessoais, ou ‘vozes textuais’ são para a comunidade discursiva para a qual estão escrevendo.

## CONSIDERAÇÕES FINAIS

Neste artigo, procurei basicamente demonstrar como os resultados obtidos através da análise contrastiva de subcorpora do ICLE, pode oferecer informações sobre características discursivas da escrita destes aprendizes que possam ser incorporadas no design de materiais pedagógicos e/ou práticas pedagógicas. Paralelamente, discuti de forma sucinta alguns dos problemas associados ao uso de corpora em certas práticas de ensino, assim como na criação de materiais pedagógicos. A esse respeito, o autor gostaria de reiterar que materiais pedagógicos devem ser influenciados *por*, mas não baseados inteira-

## LEONARDO RECSKI

mente *em* resultados obtidos a partir de corpora; a recíproca sendo verdadeira também para certos métodos de ensino.

*Agradecimentos:* Gostaria de agradecer os comentários e sugestões de dois pareceristas anônimos. Quaisquer problemas que persistam são de minha inteira responsabilidade.

## REFERÊNCIAS

ALTENBERG, Bengt. Exploring the Swedish component of the International Corpus of Learner English. In: LEWANDOWSKA-TOMASZCZYK, B. and MELIA, P.J. (Eds.). *PALC '97-Practical Applications in Language Corpora*. Lodz: Lodz University Press, 1997. p.119-132.

ASTON, Guy. Corpora in language pedagogy: matching theory and practice. In: COOK, G and SEIDHOFER, B. (Eds.). *Principle and Practice in Applied Linguistics: Studies in honor of H.G. Widdowson*. Oxford: Oxford University Press, 1995. p.257-270.

\_\_\_\_\_. Small and large corpora in language learning. In: LEWANDOWSKA-TOMASZCZYK, B. and MELIA, P.J. (Eds.). *PALC '97-Practical Applications in Language Corpora*. Lodz: Lodz University Press, 1997. p.51-62.

ASTON, Guy & BURNARD, Lou. *The BNC Handbook*. Edinburgh: Edinburgh University Press, 1998.

BIBER, Douglas. *Variations across speech and writing*. Cambridge: Cambridge University Press, 1988.

BIBER, Douglas; CONRAD, Susan & REPPEN, Randi. *Corpus Linguistics. Investigating language structure and use*.

## UTILIZANDO CORPORA DE APRENDIZES

Cambridge: CUP, 1998.

FLOWERDEW, John. Concordancing in language learning. In: PENNINGTON, M. (Ed.), *The power of CALL*, Houston, TX: Athelstan, 1996. p.97-113.

FLOWERDEW, Lynne. Integrating expert and interlanguage computer corpora findings on causality. Discoveries for teachers and students. *English for Specific Purposes*, v.17, n.4, p.329-345, 1998a.

\_\_\_\_\_. Concordancing on an expert and learner corpus in ESP. *CÆLL Journal* v.8, n.3, p.3-7, 1998b.

\_\_\_\_\_. Corpus linguistics techniques applied to text linguistics. *System* v.26, n.4, p.541-552, 2000.

GRANGER, Sylviane. New insights into the learner lexicon: A preliminary report from the international corpus of learner English. In: FLOWERDEW, L. & TONG, K.K. (Eds.), *Entering text*. Hong Kong: Hong Kong University of Science and Technology, 1993. p.102-113.

\_\_\_\_\_. (Ed.) *Learner English on Computer*. London: Longman, 1998.

GRANGER, Sylviane & RAYSON, Paul. Automatic profiling of learner texts. In: GRANGER, S. (Ed.), *Learner English on Computer*. London: Longman, 1998. p.119-131.

GRANGER, Sylviane & TRIBBLE, Chris. Learner corpus data in the foreign language classroom: Form focused instruction and data driven learning. In: GRANGER, S. (Ed.), *Learner English on Computer*. London: Longman, 1998. p.199-209.

GRANGER, Sylviane & PETCH-TYSON, Stephannie. Connector usage in the English essay writing of native and non-

LEONARDO RECSKI

native EFL speakers of English. *World Englishes*, v.15, n.1, p.17-27, 1996.

HALLIDAY, Michael & HASAN, Ruqaiya. *Cohesion in English*. London: Longman, 1976.

HOWARTH, Peter. *Advanced Writing in English as a Foreign Language: A Corpus-Based Study of Processes and Products*. Pecs: Lingua Franca Csoport, 2001.

JOHNS, Tim. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In: ODLIN, T. (Ed.). *Perspectives on Pedagogical Grammar* Cambridge: Cambridge University Press, 1994. p.293-313.

LORENZ, Gunter. Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In: GRANGER, S. (Ed.). *Learner English on Computer*. London: Longman, 1998. p.53-66.

MILTON, John. Lexical thickness and electronic gateways. Making texts accessible by novice writers. In: CADLIN, C. and HYLAND, K. (Eds.). *Writing: texts, processes and practices*. London: Longman, 1999. p.221-243.

PICKARD, Vince. Producing concordance-based self-access vocabulary package: some problems and solutions. In: FLOWERDEW, L. & TONG, K.K. (Eds.). *Entering text*. Hong Kong: Hong Kong University of Science and Technology, 1993. p.215-224.

RECSKI, Leonardo. Negociando certeza e incerteza em escrita acadêmica em inglês. *Trabalhos em Linguística Aplicada*, v.43, n.1, p.51-68, 2004.

SARDINHA, Antonio B. O corpus de aprendiz BR-ICLE. *DIRECT Papers 40*. São Paulo: LAEL – PUC/SP, 2000.

## UTILIZANDO CORPORA DE APRENDIZES

SCOTT, Mike. *WordSmith Tools*. Oxford: Oxford University Press, 1996.

STEVENS, Vance. Classroom concordancing: vocabulary materials derived from relevant, authentic text. *English for Specific Purposes*, v.10, p.10-15, 1991.

\_\_\_\_\_. Concordancing with Language Learners: Why? When? What? *CÆLL Journal*, v.6, n.2, p.2-10, 1995.

THURSTON, Jennifer & CANDLIN, Christopher. Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, v.17, n.3, p.267-280, 1998.

TRIBBLE, Chris. Concordancing and an EAP writing programme. *CÆLL Journal*, v.1, n.2, p.10-15, 1990.

\_\_\_\_\_. Some uses of electronic text in English for academic purposes. In: MILTON, J. & TONG, K.K. (Eds.). *Text Analysis in Computer Assisted Language Learning* Hong Kong: Hong Kong University of Science and Technology, 1991. p.4-14.

TURTON, Norman & HEATON, John. *Longman Dictionary of Common Errors*. London: Longman, 1996.

WIDDOWSON, Henry Context, Community, and Authentic Language. *TESOL Quarterly*, v.32, n.4, p.705-716, 1998.

WILLIS, Dave & WRIGHT, John. *Collins COBUILD Basic Grammar* London: HarperCollins, 1995.

*Recebido:* Novembro de 2004.

*Aceito:* Dezembro de 2004.

LEONARDO RECSKI

*Endereço para correspondência:*

Leonardo Juliano Recski  
Universidade Federal de Santa Catarina  
Serv. Augusto Buss, 597  
Rio Tavares  
88048-358 Florianópolis SC  
lrecski@hotmail.com