

Da Linguística de *Corpus* ao Ensino/Aprendizagem de Línguas: da teoria à prática

Sílvia Araújo
Universidade do Minho

Pedro Trábulo
Universidade do Minho

Resumo

Existe um interesse cada vez maior em *corpora* construídos a partir de repositórios de dados organizados. Cientes da relevância deste tipo de recurso no âmbito do ensino/aprendizagem de línguas, iremos apresentar, com base em exercícios que visam o ensino-aprendizagem de português por falantes nativos de inglês, as opções de pesquisa proporcionadas por alguns dos *corpora* que integram o português europeu. Como se pode ver abaixo, a partir de uma simples pesquisa num *corpus* paralelo bidirecional português-inglês (como o COMPARA), os alunos poderão facilmente aperceber-se da dimensão polissémica de um termo e da rede de termos que lhe está associada na sua língua materna.

Palavras-chave: linguística de corpus, concordância, polissemia, língua estrangeira

Abstract

There is a growing interest in corpora built from repositories of organized data. Fully aware of the relevance of this kind of resource in language teaching and learning, we will present the search options offered by some of the corpora which include European Portuguese, based on exercises aimed at the teaching and learning of Portuguese by English native speakers. As can be seen below, with a simple query in any Portuguese-English bidirectional parallel corpus (like COMPARA), the students can easily grasp the polysemy of a given word and also find the network of associated words in their mother tongue.

Keywords: corpus linguistic, polysemy, foreign language, concordance

1. Linguística de Corpus e ensino de línguas

A lexicografia foi uma das primeiras áreas da linguística a fazer uso de corpora eletrónicos. Mas graças à disponibilização crescente de corpora em diversas línguas bem como à crescente consciência da importância de dados quantitativos na investigação linguística (McEnery & Wilson 2001: 25), as abordagens baseadas em corpus ultrapassaram esses usos pioneiros de corpus pela lexicografia e têm vindo a revelar-se particularmente importantes em áreas como a tradução automática, a linguística contrastiva, a terminologia e o ensino da língua.

É deste último aspeto – o do ensino de uma língua estrangeira baseada em corpus, alicerçada na observação da ocorrência dos termos – que trata o presente estudo. Ao permitirem que as línguas possam ser estudadas com base em exemplos reais de uso e não somente por meio de textos criados com propósitos de exemplificação, os corpora funcionam como ferramentas facilitadoras da produção de material didático específico e baseado nas dificuldades reais dos alunos, contribuindo assim para a mudança na organização e nos métodos pedagógicos (Sardinha 2004).

Nos últimos anos, vários estudos e obras sobre o uso de corpora na aprendizagem de línguas têm sido editados (Granger *et al.* 2002; Aston *et al.* 2004; Aijmer 2009; Frankenberg-Garcia *et al.* 2011, Meunier 2011, entre outros). Estas publicações mostram que os corpora estão a ser usados para melhorar a aprendizagem do vocabulário (Stevens 1991; Cobb 1999), da gramática (Conrad 2001; Frankenberg-Garcia 2003; Simões & Santos 2011; Xiao & McEnery 2013; Santos 2014) e da sintaxe (McEnery & Xiao 2005; Altenberg & Granger 2012). Mas como refere Frankenberg-Garcia (2006), se há professores que aceitam os desafios e incorporam estes recursos na sala de aula, outros há que se mostram relutantes e resistem à mudança das suas práticas educativas. Apesar de não usarem diretamente corpora dentro ou fora das suas aulas, importa salientar que muitos deles fazem dos corpora um uso indireto ao usarem livros didáticos, dicionários e gramáticas concebidos por especialistas em linguística de corpora e didática das línguas.

Para uma efetiva inclusão dos corpora nas práticas docentes é essencial estabelecer um plano de formação dos professores e dos alunos. Como veremos na seção 3, seria realmente importante que os professores tentassem adaptar concordâncias (monolíngues e bilíngues), transformando-as em testes ou fichas de exercícios e que os alunos aprendessem, por sua vez, a ler essas concordâncias (Stevens 1995; Kettemann 1996; Gavioli 2001; St. John 2001; Chambers 2005, entre outros), para chegarem às regras pelas suas próprias reflexões, a partir da formulação de hipóteses. Este tipo de abordagem que Johns (1991) denomina *data-driven learning* ou DDL (e que Bernardini 2002 designa por *Corpus-aided Discovery Learning*) pode certamente ajudar a promover estratégias de intervenção que desenvolvam a autorregulação da aprendizagem, com enfoque na identificação e resolução de dificuldades, nomeadamente ao nível das atividades gramaticais.

2. Corpus: definição, tipologia e recursos existentes para o Português

Os corpora são conjuntos de textos compilados com o fim de responderem a perguntas sobre a língua ou poderem servir de treino a sistemas que a processam (Santos 2002). Como se pode ver na tabela que se segue, os principais corpora que integram o português podem ser subdivididos em corpora monolíngues ou bi/multilíngues:

CORPORA – PORTUGUÊS				
MONOLINGUES				
Português europeu				
SINCRÓNICOS	CRPC Corpus de Referência do Português Contemporâneo	CETEMPúblico Corpus Natura/ Minho (corpora jornalísticos)	CINTIL Corpus Internacional do Português	
	Lácio-Web Corpus Brasileiro (corpora do português brasileiro contemporâneo)	Português do Brasil CETENFolha (corpus jornalístico)	Br-Icle Subcorpus do ICLE: International Corpus of Learner English ¹	
DIACRÓNICOS	Corpus do Português (séc. XIV-XX)	Corpus Informatizado do Português Medieval	VERCIAL 309 obras de 55 autores portugueses publicadas entre 1500 e 1933	
BI/MULTILINGUES				
Originais + respetivas traduções para outra(s) língua(s)			Originais em duas ou mais línguas	Textos de falantes não nativos
corpora unidirecionais	corpora bidirecionais	corpora multidirecionais	corpora comparáveis	corpora de aprendentes
PER-FIDE Corpora paralelos com Pt como língua pivô	COMPARA Corpus literário (Pt-En)	OPUS 16 subcorpora especializados (7 Pt europeu; 3 Pt Brasil)	CorTec Projeto COMET (Pt Brasil)	CoMAprend Projeto COMET (5 línguas-alvo: Es, Fr, En, De, It)

Tabela 1: Corpora que integram o português

2.1. Corpora monolíngues

Os corpora monolíngues oferecem a possibilidade de estudar uma língua ao nível lexical, sintático e pragmático-semântico. Como se pode ver na tabela acima, o português europeu e do Brasil conta já com uma série de corpora monolíngues (sobretudo para a língua escrita). Apenas destacaremos aqui o serviço AC/DC (*Acesso a Corpora/Disponibilização de Corpora*)² da *Linguateca* (Santos 2014) não só pelo facto deste serviço ter vindo a disponibilizar diversos corpora gratuitamente acessíveis para a língua portuguesa, mas também por ter recentemente colocado à disposição dos

¹ - O ICLE (<http://www.uclouvain.be/en-cecl-icle.html>) reúne textos argumentativos produzidos por alunos de inglês de nível intermédio e/ou avançado de 16 nacionalidades diferentes.

² - <http://www.linguateca.pt/ACDC/>

professores de língua um sistema de criação de exercícios gramaticais sobre os corpora AC/DC. Trata-se do *Ensinador*³ (Simões & Santos 2011) que nos permite: i) gerar e filtrar uma concordância de forma a selecionarmos apenas as frases que melhor se adequam ao exercício que pretendemos criar [ver abaixo figura 1] assim como ii) visualizar o exercício e a respetiva correção [ver abaixo figura 2]:

Ensinador

A procurar “[lema=”desde|até”].” no corpus CETEMPúblico 1.7 v. 7.4 [8000 entradas aleatórias de um total de 365494 entradas.]

Selecione as concordâncias que deseja usar.

- par=ext252-clt-soc-93a-2*: Embora reafirmem a cada dois passos o carácter estritamente autárquico do acordo, os dirigentes da PE subscrevem um texto que vai **desde** a reforma do sistema político em geral até ao modelo de desenvolvimento para o país .
- par=ext417-pol-98a-1*: **Até** porque Fuzeta da Ponte, nas palavras de Alberto João, não era «grande espingarda» e «antipático», devendo por isso ter sido «demitido logo» .
- par=ext481-eco-92a-2*: Por outro lado, **até** ao final de 1991 o BNU tinha prestado 292 garantias a favor do Instituto de Apoio às Pequenas e Médias Empresas, para a liberalização prévia de incentivos concedidos por este instituto .

Figura 1: concordância gerada no *Ensinador* para *desde/até*

Ensinador

Especificação do tempo em português
Limites temporais: desde ou até?

- par=ext406-pol-91b-4*: Estão dispostos a lutar por isso, mas vão fazer uma pausa _____ 6 de Outubro .
- par=ext4269-des-92a-1*: Mas, _____ lá, faltará, decerto, muito tempo .
- par=ext5063-pol-93b-2*: O responsável explica que a sua equipa de 11 médicos operou 1200 crianças _____ o início da agressão dos radicais sérvios, em condições idênticas às do resto do hospital .

Ver Solução Descarregar: enunciado Adicionar título: Especificação do tempo em portugú

Figura 2: visualização do exercício resultante da concordância anterior

2.2. Corpora multilingues

Também existem *corpora* que reúnem textos de duas (bilingues), três ou mais línguas (multilingues). Estes *corpora* dividem-se em paralelos e comparáveis (Granger 2010). Como referem McEnery e Xiao (2007: 139), os corpora paralelos caracterizam-se por alinhar originais (= língua fonte, LF) e suas respetivas traduções (= língua alvo, LA). Estes corpora podem ser: i) unidirecionais (LF → LA), ii) bidirecionais (LF ↔ LA) ou iii) multidirecionais (LF → [LA₁, LA₂, LA₃, LA₄, etc]). No domínio literário, a *Linguateca*

³ - <http://www.linguateca.pt/Ensinador/>

desenvolveu o COMPARA, que é um corpus literário bidirecional português-inglês.

Para ampliar este trabalho de compilação não só a outros pares de línguas, mas também a uma maior variedade de textos, o projeto *Per-Fide*⁴ (Araújo *et al.* 2010; Almeida 2014) tem vindo a criar, desde abril de 2010, um corpus paralelo multilíngue composto por um conjunto de subcorpora pertencentes a diversos domínios e.g. religioso, literário, jornalístico e técnico, que envolve um total de sete línguas (Português, Español, Russian, Français, Italiano, Deutsch, English). Os corpora paralelos compilados contêm textos originais nas sete línguas a par das respetivas traduções no maior número possível de línguas.

Para além de oferecer, tal como o corpus *Per-Fide*, um leque muito variado de línguas, o *Opus*⁵ oferece-nos ainda a possibilidade de realizar pesquisas que permitem visualizar duas, três ou mais línguas numa única concordância. Este corpus multilíngue disponibiliza, de forma gratuita, 16 subcorpora especializados: textos legislativos (Parlamento Europeu); textos de economia e finanças (Banco Central Europeu); textos do domínio da saúde (Agência Europeia de Medicina); manuais técnicos de informática; corpus de legendas. O português europeu é representado em 7 desses subcorpora e o do Brasil em apenas 3.

2.3. Corpora de aprendentes

A maior parte dos corpora de aprendentes é de linguagem escrita e integra produções de inglês língua estrangeira. De facto, dentro dos 127 ‘learner corpora’ que aparecem elencados na página *Learner corpora around the world*⁶ da Universidade Católica de Lovaina, 74 deles são dedicados ao inglês. Ao incluir produções redigidas por alunos brasileiros em cinco línguas alvo (alemão, espanhol, francês, inglês e italiano), o *Corpus Multilíngue de Aprendizizes* (CoMAprend)⁷ compilado na Universidade de São Paulo destaca-se precisamente por criar um ‘corpus de aprendizizes’ multilíngue. Em Portugal, importa destacar os dois projetos que se seguem:

- o projeto de *Recolha de dados de aprendizagem de português língua estrangeira* realizado ao abrigo de um protocolo entre o Instituto Camões e o Centro de Linguística da Universidade de Lisboa reúne 470 produções escritas, realizadas por 397 falantes de 28 línguas maternas diferentes.

⁴ - Acessível gratuitamente em: <http://per-fide.ilch.uminho.pt/query>, o Corpus *Per-Fide* apresenta um recurso complementar à pesquisa em corpora, os PTDs – Dicionários Probabilísticos de Tradução – gerados automaticamente (Almeida *et al.* 2014). Ao gerar um PTD, o sistema de pesquisa apresenta uma família paradigmática de equivalentes com a sua respetiva percentagem de correspondência direta entre o termo de partida e os vários termos de chegada possíveis dentro do corpus selecionado pelo utilizador.

⁵ - O corpus está disponível gratuitamente em: <http://opus.lingfil.uu.se/bin/opuscqp.pl>

⁶ - <http://www.uclouvain.be/en-cecl-lcworld.html>

⁷ - O corpus está disponível em: <http://www.fflch.usp.br/dlm/comet/comaprend.html>

- Os materiais do *Corpus de Produções Escritas de Aprendentes de PL2 (PEAPL2)* da Universidade de Coimbra foram produzidos por 391 falantes de 39 línguas maternas diferentes.

As produções escritas recolhidas no âmbito destes dois projetos podem ser descarregadas⁸ gratuitamente. Este tipo de projetos tem como principal objetivo recolher produções de aprendentes não nativos com vista à criação de uma base de dados que possa apoiar a investigação em aquisição/aprendizagem de línguas estrangeiras (Aston 2004; Granger 2002; 2012) e, de modo particular, a formação de professores e a produção de materiais didáticos.

3. Como pesquisar em corpora

Importa referir que um corpus permite pesquisas mais avançadas se estiver etiquetado e lematizado (Frankenberg-Garcia, 2004; Santos 2012). Poder-se-á, neste caso, proceder a pesquisas de expressões que implicam um certo grau de variabilidade morfossintática (flexões, inserção de modificadores, etc.). Por exemplo, para procurarmos colocações do tipo "tomar + NOME" no corpus jornalístico do CETEMPúblico, é necessário introduzir a fórmula que se segue: **[word="tomar"] [pos="N"]** (i.e. uma fórmula do tipo: [atributo="valor"]⁹).

Para perceber a sintaxe de pesquisa acima transcrita, convém saber que:

- o atributo *word* permite a pesquisa de uma **palavra (forma)**:
→ **[word="tomar"]** (esta fórmula procura apenas formas com a forma ortográfica *tomar*)
- o uso do atributo **lema** permite a pesquisa das palavras pela sua **forma base**:
→ **[lema="tomar"]** (neste caso, obtemos formas que têm *tomar* como lema, i.e. todas as formas flexionadas de *tomar* tais como *tomou*, *tomaste*, *tomaram*, etc.).
- a pesquisa de uma palavra com determinada categoria morfossintática é feita através do atributo **pos** (*part-of-speech*):
→ **[word="tomar"] [pos="N"]** devolve lexias que correspondem à palavra *tomar* seguida de nomes. Obtemos então uma família paradigmática de construções com verbo-suporte (*tomar conta (da loja/da casa/de mim)*,

⁸ - Os dados destes projetos estão respetivamente disponíveis em: <http://www.clul.ul.pt/pt/recursos/314-corpora-of-ple>; <http://www.uc.pt/fluc/rcpl2/dados>.

⁹ - Informação adicional sobre anotação está disponível para consulta em: http://per-fide.ilch.uminho.pt/docs/summerschool/how_to_query_corpora.pdf; http://per-fide.ilch.uminho.pt/docs/summerschool/Idaete_Dias_et_al/exercises.pdf.

conhecimento, consciência, parte (na galhofa/no motim/na expedição), partido, posse, posição, nota (dos compromissos/de tudo/), medidas, decisões, contacto, ...) cujo semantismo é suportado pela base nominal.

Através da combinação das diferentes opções de pesquisa descritas acima

word: palavra (forma) → [word="..."]
lema: forma base da palavra → [lema="..."]
pos: (part-of-speech) categoria grammatical → [pos="..."]

é possível construir pesquisas avançadas como as que são de seguida exemplificadas:

. [lema="andar"][word="a"][pos="V.*"]
→ *andar a trabalhar/ando a escrever/andam a pedir/ (= perífrase andar a + Infinitivo).*

Também é possível recorrer a operadores de repetição ({n} exactamente n vezes; {n,} n ou mais vezes; {,n} até n vezes) ou a expressões combinadas (| alternância; () junção; ! negação) para proceder a consultas mais refinadas:

. [pos="DETard"]|[pos!="N"]{2,3}[pos="ADJ"] dá como resultado sequências com um artigo definido seguido por 2 ou 3 formas que não sejam nomes comuns, seguidas por um adjetivo → *o mais depressa possível; o mais cedo possível; o que é mais característico; o que pode parecer incongruente; o que é pior; o que é perfeitamente compreensível;*

. [pos="N|ADJ"]{3,} dá como resultado sequências que tenham pelo menos 3 adjetivos e nomes comuns consecutivos (em qualquer uma das ordens possíveis)

→ *tinta verde-clara viscosa, estilo gótico vitoriano, cabelo preto ondulado, défice comercial britânico, família tipo maria-vai-com-as-outras, articulação reumática sensível, minúsculo comprimido verde-claro, atrasado mental hilariante, grande crise económica, ...*

4. Pesquisa em corpus: exercícios para ensino do PLE

Para dar conta das potencialidades pedagógicas de um corpus, iremos apresentar alguns exercícios que incidem sobre o verbo *chegar* em português e seus equivalentes em inglês para os níveis de proficiência A2/B1.

4.1. Verbetes tradicionais não baseados em corpus

Importa salientar, desde já, que o léxico é frequentemente apresentado de forma totalmente descontextualizada nos dicionários não baseados em corpus:

Priberam http://www.priberam.pt/DLPO/chegar	Infopédia http://www.infopedia.pt/pesquisa-global/chegar
chegar – Conjugar (origem duvidosa) verbo intransitivo 1. Vir. 2. Dar entrada em. 3. Atingir, alcançar, durar até. 4. Ir, prolongar-se, ir ter, ir dar. 5. Tocar. 6. Aproximar-se. 7. Ascender, importar. 8. Bastar. 9. Ir até ao ponto de. 10. Subir até. 11. Acercar-se. 12. Começar. 13. Conseguir. 14. Ser bastante alto para. 15. Dar pancadas em, bater, surrar. verbo transitivo 16. Aproximar, mover para perto. 17. Ser suficiente para algo ou alguém (ex.: esta comida não chega para tantos convidados). = DAR verbo pronominal 18. Aproximar-se; vir ter a.	chegar – conjugação verbo transitivo e intransitivo 1. vir (de algum lugar) 2. colocar(-se) perto; aproximar(-se) 3. ser suficiente (para); bastar verbo intransitivo 1. dar entrada em 2. regressar 3. ter início; começar verbo transitivo 1. attingir (determinado lugar ou ponto); alcançar 2. dar (algo) a; passar 3. pôr em contacto 4. igualar-se (a); comparar-se (a) 5. sovar; bater 6. levar a fêmea à cobrição verbo pronominal 1. aproximar-se 2. resolver-se [...]

Tabela 2: verbetes de *chegar* nos dicionários online *Priberam* e *Infopédia*

Os dois verbetes acima mencionados que retirámos dos dicionários *Priberam* e *Infopédia* da Língua Portuguesa são de muito pouca utilidade para alunos que estejam ainda numa fase inicial de aprendizagem do português, pois os diferentes valores semânticos de *chegar* vêm desprovidos, como vemos, de qualquer exemplo. Nesse sentido, parece-nos que o “mergulho” nos dados autênticos de um *corpus* pode complementar a pesquisa realizada a partir de outros recursos mais convencionais (tais como os dicionários ou ainda os tradutores automáticos), pois fornece as ocorrências do termo pesquisado nos seus diferentes contextos.

4.2. O verbo *chegar* e seus equivalentes em inglês

É hoje praticamente consensual que a significação de uma forma linguística não existe por si mesma: define-se através dos variados modos de relação com o cotexto (também “contexte imédiat”, segundo Franckel 2006) em que se inscreve. A dinâmica interna à própria linguagem é, por conseguinte, da ordem da interação. Os exercícios que propomos abaixo têm precisamente como objetivo ajudar os estudantes de PLE a apreenderem, com base numa pesquisa

em corpus, a dimensão polissêmica de um termo (neste caso, o verbo *chegar*) e a rede de termos que lhe está associada na sua língua materna.

4.2.1. *Chegar* como verbo pleno

O primeiro exercício abaixo transcrito parte deliberadamente da língua materna do aluno (neste caso, o inglês) de forma a dar conta da relação de equivalência entre *arrive*-*chegar*:

1. Clique em *Pesquisa Simples* no COMPARA. Escreva *arrive* na caixa de pesquisa de inglês para português e clique no botão Pesquisar (de inglês para português).
- 1.1. Quantas ocorrências de *arrive* existem?
- 1.2. Percorra os resultados para determinar qual é o verbo português que traduz mais frequentemente o verbo *arrive*?

Ou seja, com este exercício, pretende-se mostrar que a tradução de *arrive* em português não parece suscitar grandes dificuldades, dado que se trata, quase sempre, de uma tradução que mobiliza o equivalente português mais direto de *arrive*, i.e. *chegar*:

Procura: *arrive* Pedido de: **concordância em contexto** Direção da pesquisa: **De inglês para português** Resultados: **78** ocorrências · Expressão de pesquisa: "**arrive**"

Descrição do corpus usado nesta procura: 1435826 palavras portuguesas, 1542762 palavras inglesas, 97723 unidades de alinhamento.

Concordância

EBDL5(1):	Nightly descending through the baroque cloud That decorates these hills, riding on air, Thousands arrive by dream at their desire.	Descendo à noite pela nuvem branca Que ornamenta estes montes, rasgando o ar, Em sonhos chegam milhares ao que desejam.
EBDL5(99):	"The customers are confused, anxious, when they arrive here.	-- Os clientes estão confusos e ansiosos quando aqui chegam ,
EBDL5(1095):	'And who knows whether the shops will be open when we arrive ?' -- E sabe-se lá se as lojas vão estar abertas quando chegarmos!	
EBDL6(43):	Many conferees arrive at their destinations this summer more than usually fatigued, dehydrated and harassed; the dying fall of the engines' <i>Whhhheeeeeeeeeeeeeee</i> , as the power is finally switched off, is sweet music to their ears, but their chatter is undiminished, their demand for information insatiable.	Neste Verão são muitos os conferencistas que chegam aos seus destinos mais cansados, desidratados e molestados do que é habitual, mas o moribundo zunido dos motores, <i>UUUUUUuuuuuu</i> , quando o aparelho finalmente se imobiliza, é uma doce melodia para os seus ouvidos, ainda que a tagarelice não diminua e a sua ânsia por informações seja insaciável.
EBDL6(877):	He was expected at the Vienna conference on Narrative last week, but he did not arrive .	Era esperado no de Viena, sobre Narrativa, a semana passada, mas não apareceu .
EBDL6(1632):	«He was expected, but did not arrive , for some reason.	-- Estava-se à espera dele, mas, por qualquer motivo, não apareceu .

Figura 3: Concordância bilingue de *arrive* no corpus paralelo COMPARA

Este tipo de pesquisa fornece ocorrências do termo pesquisado nos seus diferentes contextos – a esta lista de ocorrência atribui-se o nome de concordância.

2. Clique novamente em Pesquisa simples. Escreva *chegar* na caixa de pesquisa de português para inglês e clique no botão Pesquisar (de português para inglês).
- 2.1. Quantas ocorrências de *chegar* existem?
- 2.2. Quais as diferenças entre os resultados obtidos para *arrive* e os resultados para *chegar*?

Depois de ter observado, aquando do primeiro exercício, que *arrive* ocorre 78 vezes na direção linguística de inglês para português, o aluno conclui que *chegar* surge com muito mais frequência na direção de português para inglês,

pois obtemos, neste caso, 428 ocorrências de *chegar*. A disparidade destes resultados aponta para a polissemia do verbo *chegar* (Pauly 2010) que parece ter desenvolvido usos mais abstratos a partir do seu uso mais concreto que denota um movimento espacial: *chegar a algum lugar* (Ferreira 2011). Por conseguinte, a determinação das suas diferentes significações passa necessariamente, em inglês, por verbos diferentes (*to come, reach, approach, para além de to arrive*), como tentamos mostrar a seguir com (cinco) dos contextos mais característicos de *chegar* extraídos do COMPARA.

3. Apresenta-se a seguir um excerto da concordância bilingue de *chegar* para dar conta dos principais equivalentes deste verbo em inglês.

EXEMPLO EM PORTUGUÊS	EXEMPLO EM INGLÊS	EQUIVALENTE EM INGLÊS
1. Finalmente <i>chegámos</i> .	We've finally <i>arrived</i> .	A: to arrive
2. O comboio <i>está a chegar</i> a Euston.	The train <i>is approaching</i> Euston.	B: to approach
3. Talvez tenha <i>chegado</i> naturalmente ao fim dos seus dias.	Perhaps it's <i>reached</i> the end of its natural life.	C: to reach
4. Levantou-se para lhe <i>chegar</i> a tigela com a papa de milho.	She got up to <i>take</i> him his bowl of mealie porridge.	D: to give
5. Já <i>chega</i> .	That's <i>enough</i> .	E: to be enough

Neste exercício seguinte, fornecemos aos alunos outras 10 ocorrências de *chegar* que encontrámos na concordância bilingue de *chegar-arrive* e pedimos-lhes que procurem determinar a significação de cada uma destas ocorrências de acordo com os 5 equivalentes ingleses indicados em 3:

3.1. Indique o valor semântico de *chegar* em cada um dos enunciados que se seguem, recorrendo aos cinco equivalentes ingleses (de A a E) apresentados acima:

	Exemplos retirados do corpus COMPARA	Equivalente (em inglês)
1	Escalo, <i>chego</i> ao topo e respiro ar limpo.	
2	As pessoas já estavam a começar a <i>chegar</i> para o concerto da noite.	
3	O velho <i>foi chegando</i> , vagaroso como era seu costume.	
4	Tu tens feijão em casa, que <i>chegue</i> para nós?	
5	Eu informo-o quando o médico <i>chegar</i> .	
6	Quando estávamos a <i>chegar</i> ao portão da frente calámo-nos.	
7	Já dormiu que <i>chegue!</i>	
8	Como ela estava sentada, o lençol não <i>chegava</i> até lá acima.	
9	<i>Chegue-me</i> a água, por favor.	
10	Afasta-te, não <i>te chegues</i> a mim, posso contagiar-te.	

4.2.2. *Chegar* como verbo auxiliar

Nesta segunda etapa, trata-se de fornecer aos alunos a fórmula que lhes permita aceder à concordância de *chegar* como verbo auxiliar na perífrase *chegar + a + INF*:

3.2. Utilize novamente a *Pesquisa Simples*. Escreva a sintaxe seguinte: [lema="chegar"][word="a"][pos="V"] na caixa de pesquisa de português para inglês e clique no botão Pesquisar (de português para inglês).

3.2.1. Esta pesquisa mostra que *chegar* pode aparecer em construções do tipo *chegar + a + INF*. Quantas ocorrências encontrou para este tipo de construção?

A pesquisa avançada por restrição de alinhamento permite fazer uma pesquisa nas duas línguas do corpus simultaneamente. Este tipo de pesquisa pode ter utilidade para saber, por exemplo, se uma palavra/estrutura na LF corresponde a outra palavra/estrutura na LA:

3.2.2. Clique na hiperligação da *Pesquisa Avançada*. Selecciona a direção de português para inglês. Escreva "even" na caixa das restrições de alinhamento do lado direito e "[lema="chegar"][word="a"][pos="V.*"]" na caixa de pesquisa da esquerda. Clique em Procurar.

3.2.2.1. Quantas ocorrências de *even* existem no lado inglês do corpus?

Assim, ao escrevermos a fórmula [lema="chegar"][word="a"][pos="V.*"] na caixa de pesquisa do lado esquerdo e *even* na caixa de restrição de alinhamento [ver abaixo figura 4], só serão mostrados os resultados contendo a construção *chegar a Inf.* no lado português e *even* no lado inglês da concordância [ver abaixo figura 5]:

The screenshot shows the 'Pesquisa avançada' (Advanced Search) interface. On the left is a navigation menu with options like 'Início', 'Pesquisa' (with sub-options 'Simples', 'Avançada', 'Ultra-avançada'), 'Atuda', 'Textos do COMPARA', 'Informações gerais', 'Documentação específica', and 'Linguateca'. Below the menu are 'Notícias' and 'Restrições de alinhamento' links. The main area is titled 'Pesquisa avançada' and contains a description of advanced search capabilities. A progress bar shows four steps: 1. Seleccione direcção da pesquisa, 2. Introduza expressão, 3. Escolha partes específicas do corpus (opcional), and 4. Especifique os resultados. The current step is 2.1 'Pesquisa por palavra ou expressão', where the search term '[lema="chegar"][word="a"][pos="V.*"]' is entered. Step 2.2 'Mais opções de pesquisa' is also active, showing the alignment restriction 'even' in the right-hand box. There are checkboxes for 'Não fazer distinção entre maiúsculas e minúsculas' and 'Juntar acentos e cedilhas automaticamente' for both search and alignment boxes. At the bottom are 'Procurar' and 'Limpar todo o formulário' buttons.

Figura 4: Pesquisa Avançada (restrição de alinhamento) no COMPARA

Procura: [lema="chegar"][word="a"][pos="V.*"] Pedido de: concordância em contexto Direção da pesquisa: De português para inglês Resultados: 39 ocorrências. Expressão de pesquisa: [lema="chegar"][word="a"][pos="V.*"] :COMPARA_ING_ANOT "even"

Descrição do corpus usado nesta procura: 1435926 palavras portuguesas, 1542762 palavras inglesas, 97723 unidades de alinhamento.

Concordância

EBDI_1T1 (1690):	Cheguei a forçar o riso em relação a certos diálogos que escrevera – algo que habitualmente não faço.	I even forced a laugh at some of my own lines, something I never normally do.
EBJT1(687):	Ela chegara a fazê-lo sentir que o dinheiro não tinha importância, que não estava a perder o controlo da sua capacidade para gerir as finanças da propriedade.	She'd even made him feel that the money didn't matter, that his seeming ability to manage the farm's finances was not running away with him.
EBMS1(185):	Lágrimas de felicidade de novo me escorriam nas faces e cheguei a levantar os olhos rasos de lágrimas para o sol bendito que tanta alegria me dava.	Soft tears again bedewed my cheeks, and I even raised my humid eyes with thankfulness towards the blessed sun which bestowed such joy upon me.

Figura 5: Resultado da concordância da pesquisa por restrição de alinhamento

Esta construção perifrástica com *chegar* parece nutrir uma relação privilegiada com *even* e com outras construções do tipo [lema="get"][word="to"][pos="V.*"] ou ainda [lema="begin"][word="to"][pos="V.*"]. A partir destes diferentes equivalentes ingleses, os aprendentes poderão mais facilmente emitir hipóteses quanto ao sentido de *chegar* neste contexto:

3.2.2.2. Que valor(es) semântico(s) atribui a esta construção perifrástica inglesa que aparece neste contexto?

3.2.2.3. Recorra a este tipo de pesquisa (por restrição de alinhamento) para recolher, pelo menos, mais uma das traduções possíveis desta construção perifrástica em inglês?

4.2.3. *Chegar* em unidades fraseológicas (colocações, expressões idiomáticas, etc.)

Nesta terceira e última etapa que se prende com o uso de *chegar* em expressões mais ou menos fixas, o COMPARA pode funcionar, uma vez mais, como um dicionário bilingue contextualizado que permite pesquisas multipalavras. O professor poderá então propor uma série de lexias complexas construídas com o verbo *chegar* e encorajar os alunos a fazer uma pesquisa num corpus paralelo para descobrirem se estas lexias têm uma correspondência direta ou não em inglês:

4. Retroceda até à *Pesquisa Simples* para pesquisar as 4 expressões que se seguem:

- A. "chegar" "a" "uma" "conclusão"
- B. "chega" "de" "(conversa|paleio|treta|palavreado)"
- C. "não" "chegar" "aos" "calcanhares" "de"
- D. [lema="chegar"][word="a"][word="mostarda"][word="ao"][word="nariz"]

4.1. Complete a tabela abaixo indicando quantas ocorrências encontrou para cada uma delas.

4.2. Para ilustrar cada uma dessas expressões, transcreva, nessa mesma tabela, uma das ocorrências que encontrou na parte do português e sua respetiva tradução inglesa, como no exemplo.

Exemplos retirados do COMPARA			
	Oc.	PT	EN
A	5	[...] e <i>cheguei</i> a uma conclusão única.	[...] and I <i>have arrived</i> at one conclusion.
B			
C			
D.			

5. Considerações finais

Embora alguns autores, como Gellerstam (1996), alertem para os efeitos potencialmente nefastos de se expor os aprendentes de uma língua estrangeira às idiossincrasias típicas das traduções, procurámos mostrar, nesta comunicação, que o uso de *corpora* paralelos pode ajudar no desenvolvimento (meta)linguístico dos alunos. De facto, ao aprender a “ler” e a “analisar” concordâncias bilíngues, os alunos são necessariamente incitados a:

- . formular hipóteses e a compará-las com as dos colegas,
- . estabelecer relações e comparações intra e interlinguísticas e, a partir delas, induzir regras gramaticais que reforçam o diálogo entre a língua materna e a língua estrangeira.

Deste modo, a pesquisa em *corpus* permite que a abordagem da gramática possa ser indutiva, orientando os alunos na descoberta autónoma do funcionamento da língua. A exploração de *corpora* permite-nos dar conta, por exemplo, da polissemia dos verbos e da forte assimetria sintático-semântica que existe, na passagem de uma língua para a outra, entre verbos aparentemente equivalentes. A partir das concordâncias resultantes de uma pesquisa bidirecional efetuada no COMPARA, é possível elaborar, como vimos, exercícios que ajudam os alunos de PLE a apreender os diferentes sentidos que um determinado verbo adquire em contexto através da rede de equivalentes por eles mobilizada na sua língua materna.

6. Referências bibliográficas

- Almeida, José João; Araújo, Sílvia; Carvalho, Nuno; Dias, Idalete; Oliveira, Ana; Santos, André; Simões, Alberto (2014): “The Per-Fide Corpus: a new resource for corpus-based terminology, contrastive linguistics and translation studies”. In Tony Berber Sardinha & Telma São Bento Ferreira (eds.) *Working with Portuguese Corpora*. Bloomsbury Academic.177-200.
- Aijmer, Karin (2009): *Corpora and Language Teaching*. Amsterdam: John Benjamins.
- Altenberg, Bengt; Granger, Sylviane (2012): “The Grammatical and Lexical Patterning of MAKE in Native and Non-Native Student Writing”. In: Doug Biber & Randi

- Reppen, *Corpus Linguistics Volume 4: Methods and Applications*. Sage: London. 219-238.
- Araújo, Sílvia, Alberto Simões, José João Almeida, & Idalete Dias (2010): “Apresentação do projeto *Per-Fide*: Paralelizando o Português com seis outras línguas”. *Linguamática*, Vol.2, Nº 2. 71-74.
- Aston, Guy, Bernardini, Sílvia & Stewart, Dominic (eds.) (2004): *Corpora and Language Learners*. Amsterdam: John Benjamins.
- Bernardini, Sílvia (2002): “Exploring new directions for discovery learning”. In B. Kettemann & G. Marko (Eds.) *Teaching and learning by doing corpus analysis*. New York: The Edwin Mellen Press. 165-182.
- Chambers, Angela (2005): “Integrating Corpus Consultation in Language Studies”. *Language Learning & Technology*, 9(2). 111-125
- Cobb, Tom (1999): “Breadth and depth of vocabulary acquisition with hands-on concordancing”. *Computer Assisted Language Learning* 12. 345 - 360.
- Conrad, Susan (2000): “Will corpus linguistics revolutionize grammar teaching in the 21st century?”. *TESOL Quarterly* 34. 548–60.
- Ferreira, Ediene (2011): “Metáfora e gramaticalização: um estudo do verbo chegar”. *Veredas* 2/2011. 168-178.
- Franckel, Jean-Jacques (2006): “Situation, contexte et valeur référentielle”. *Pratiques* n° 129/130: 51-70.
- Frankenberg-Garcia, Ana, Flowerdew, L., Aston, G. (2011): *New Trends in Corpora and Language Learning*. London: Continuum.
- Frankenberg-Garcia, Ana (2003): “A Construção (e alguns usos) do corpus Compara”. In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho de 2003), Braga : Universidade do Minho, pp. 15-23.
- Frankenberg-Garcia, Ana (2004). *COMPARA - Aula Prática em Português*. 28 de Setembro de 2004. <http://www.linguateca.pt/COMPARA/docum/AulaPratica.pdf>
- Frankenberg-Garcia, Ana (2006): “Raising teachers' awareness to corpora”. In *Seventh International Conference on Teaching and Language Corpora (TALC7)* (Paris, França, 1-4 de Julho de 2006). 1-3.
- Gavioli, Laura (2001): “The learner as researcher: introducing corpus concordancing in the classroom”. In: Aston, G. (ed.) *Learning with corpora*. Houston TX: Athelstan, 108-137.
- Gellerstam, Martin (1996): “Translations as a source for cross-linguistic studies”. In Karin Aijmer, Bengt Altenberg e Mats Johansson (eds.) *Languages in contrast: papers from a symposium on text-based crosslinguistic studies*. Lund Studies in English 88. Lund University Press. 53-62.
- Granger, Sylviane (2012): “How to use foreign and second language learner corpora?”. In: Mackey, A. & Gass, S.G., *A Guide to Research Methods in Second Language Acquisition*. Basil Blackwell: Chichester. 7-29.
- Granger, Sylviane (2010): “Comparable and translation corpora in cross-linguistic research. Design, analysis and applications”. *Journal of Shanghai Jiaotong University*, Vol. 2. 14-21.
- Granger, Sylviane, Petch-Tyson, Stephanie & Hung, Joseph (eds) (2002): *Computer Learner Corpora, Second Language Acquisition and Foreign Language*

- Teaching*, éd. (*Language Learning and Language Teaching*; 6). Benjamins: Amsterdam and Philadelphia.
- Johns, Tim (1991): "Should you be persuaded": two samples of data-driven learning materials. In T. Johns and P. King (eds.) *Classroom Concordancing ELR Journal* 4. University of Birmingham.
- Kettemann, Bernhard (1996): "Concordancing in English Language Teaching". In S. Botley, J. Glass, T. McEnery and A. Wilson (eds.) *Proceedings of Teaching and Language Corpora*. 4-16. Lancaster University.
- McEnery, Tony and Richard Xiao (2007): "Parallel and comparable corpora: The state of play". In Y. Kawaguchi, T. Takagaki, N. Tomimori and Y. Tsuruga (eds.) *Corpus-Based Perspectives in Linguistics*. Amsterdam: John Benjamins. 131-145.
- McEnery, Anthony; Xiao, Zhonghua (2005): "Help or help to: What do corpora have to say?". *English Studies* 86(2). 161-187.
- McEnery, Tony & Wilson, Andrew (2001): *Corpus Linguistics. An Introduction* (2nd edition). Edinburgh: Edinburgh University Press.
- Meunier, Fanny (2011): "Corpus linguistics and second/foreign language learning: exploring multiple paths". *Revista Brasileira de Linguística Aplicada*. v. 11, n. 2. 459-477.
- Pauly, Émilie (2010): *La polysémie. Réflexion théorique, méthodologique et application à la lexicographie. L'exemple des verbes aller, partir et tirer en français contemporain*. Paris: L'Harmattan.
- Santos, Diana (2014): *First steps of Gramateca: a corpus-based grammar initiative for Portuguese, driven by Linguateca*, presentation at the Corpus Linguistics Group, University of Oslo, 20 February 2014.
- Santos, Diana (2014): "Corpora at Linguateca". In Tony Berber Sardinha & Telma São Bento Ferreira (eds.) *Working with Portuguese corpora*. Bloomsbury. 219-236.
- Sardinha, Tony Berber (2004): *Linguística de Corpus*. Editora Manole: Barueri S.P.
- Simões, Alberto & Diana Santos (2011): *Ensinador: corpus-based Portuguese grammar exercises*. *Procesamiento del Lenguaje Natural* 47: 301-309.
- St. John, Elke (2001): "A case for using a parallel corpus and concordancer for beginners of a foreign language". *Language Learning & Technology*, 5(3). 185-203.
- Stevens, Vance (1995). "Concordancing with Language Learners: Why? When? What?" *CAELL journal*, 6 (2): 2-10.
- Stevens, Vance (1991): "Classroom concordancing: vocabulary materials derived from relevant, authentic text". *English for Specific Purposes* 10: 35-46.
- Xiao, Richard & McEnery, Tony (2013): "Grammar and corpora". In *The Encyclopedia of Applied Linguistics*. Malden and Oxford: Wiley-Blackwell.