

APRESENTAÇÃO DOS RECURSOS LINGUÍSTICOS PARA A LÍNGUA PORTUGUESA E PROCEDIMENTOS DA ANÁLISE DE CORPUS DO LÉXICO PORTUGUÊS

Introdução

Observa-se, actualmente, em todo o mundo, um interesse crescente pela criação de recursos linguísticos, nomeadamente *corpora* e léxicos de grandes dimensões, o que tem sido possível graças ao extraordinário desenvolvimento da informática e do poder dos computadores. Estes recursos linguísticos específicos de cada língua, em associação com tecnologias adequadas à extracção de dados e de conhecimentos, constituem pré-requisitos indispensáveis a um grande conjunto de trabalhos de investigação. Os *corpora* proporcionam novas maneiras de estudar as línguas, das quais resultam descrições, generalizações e hipóteses teóricas de grande consistência porque são fundamentadas nos dados empíricos.

Os *corpora* da língua portuguesa que apresentamos ao seguir, léxicos e dicionários monolíngues, bilingues e plurilingues gerais e especializados electrónicos que são disponíveis on-line, facilitam novas descrições do português baseadas em dados reais, oferecem possibilidades para estudos contrastivos entre variedades do português, visando estabelecer factores de unidade e de diversidade, proporcionam materiais para estudos contrastivos entre línguas e também materiais para o ensino da língua.

1 Apresentação do projecto Linguateca

No presente artigo dedicamo-nos à apresentação dos recursos linguísticos para o estudo e processamento da língua portuguesa dentro das mais modernas tendências da linguística de corpus. Apresentamos o projecto AC/DC (Acesso a *corpora* /Disponibilização de *corpora*) cujos principais objectivos são disponibilizar todos os *corpora* a partir de um único local, melhorar a informação associada a esses *corpora* e desenvolver uma interface fácil de utilizar. Apresentamos o projecto Linguateca, cujos objectivos são: melhorar significativamente as condições para o processamento do português, nomeadamente facilitar a disponibilização dos recursos existentes, ajudar ao desenvolvimento de recursos públicos, fornecer programas para obter *corpora* através da Internet, criar *corpora* suficientemente grandes que possam ser usados como referência e facilitar o processamento de *corpora* portugueses em geral.

O trabalho da Linguateca pode ser descrito através do modelo IRA (Informação - Recursos - Avaliação):

- Informação: é um portal constantemente actualizado sobre o processamento computacional da língua portuguesa, contendo catálogo de recursos entre os quais encontramos *corpora*, léxicos e dicionários, enciclopédias, material didáctico, textos em português. O portal contém também a lista dos actores que são diferentes grupos, centros, institutos e projectos, seja europeus seja internacionais, associações e instituições, empresas, páginas pessoais, revistas sobre o português. O portal da Linguateca fornece também catálogo de publicações sobre o processamento computacional

da língua portuguesa, um fórum sobre assuntos relacionados com o processamento computacional da língua portuguesa, contendo notícias, ofertas de emprego e anúncios de conferências e de cursos, um repositório na rede de teses e outros trabalhos na área e um sistema de procura sobre a área do processamento computacional do português, construído para facilitar a navegação no portal.

- Recursos: os criadores do projecto Linguateca criam e dão acesso através da rede a um conjunto considerável de recursos para a engenharia da linguagem em português, assim como serviços que disponibilizam recursos. O projecto AC/DC (Acesso a corpora/Disponibilização de corpora) tem como objectivo tornar o acesso a corpora em português mais fácil.

O catálogo de recursos oferece um lista numerosa de recursos para a língua portuguesa, no primeiro lugar têm os corpora, alguns deles anotados, que vão ser apresentados particularmente. No segundo lugar encontramos dicionários e léxicos.

Dicionários e recursos léxicos do projecto Linguateca:

Esta lista agrupa todos os locais onde se encontra a informação sobre palavras portuguesas. Os recursos linguísticos estão divididos em dois grupos, o primeiro é o local dos léxicos gerais, onde encontramos vários dicionários monolíngues, bilingues e multilingues, o segundo local disponibiliza, salvo os dicionários monolíngues, bilingues e multilingues também glossários, listas de termos e listas de frequências.

Enciclopédias e Ontologias, Gramática e Material Didáctico:

Linguateca oferece também Enciclopédias (11 endereços) e Ontologias (9 endereços) electrónicas. No capítulo Ontologias encontramos vários Thesaurus, por exemplo Thesaurus Cadeia Alimentícia, no qual encontramos uma descrição pormenorizada dos frutos brasileiros e o Thesaurus de Folclore e Cultura Popular Brasileira. Este tipo de manuais são, na nossa opinião, embora seja só a minoria deles acessível on-line, uma excelente ferramenta para o trabalho de tradutores. Na página Gramática a Linguateca actualmente já não oferece endereços. Na página Material Didáctico a Linguateca disponibiliza o acesso a 50 endereços electrónicos, 22 do português para lusófonos, 26 do português para estrangeiros. Também oferece 2 endereços de Cursos de literatura. Linguateca fornece também a página electrónica de Textos em português contendo 54 endereços ou seja 31 endereços de textos literários, 10 endereços de textos legais, 11 endereços de outros textos e 2 endereços de textos multilingues. Nas páginas da Linguateca encontramos também os endereços que disponibilizam Comunicação social, ou seja 114 sites de Jornais e revistas de Brasil, Jornais e Revistas de Portugal, Jornais e Revistas em português de outras partes do mundo, 16 locais de Rádios e 6 de Televisões. Linguateca também oferece dois endereços importantes, <http://www.bn.br/> in <http://www.bn.pt/>, nos quais atingimos a Fundação da Biblioteca Nacional do Brasil e a Biblioteca Nacional de Portugal.

2 Corpora da língua portuguesa

Na página da Linguateca que fornece os endereços dos corpora da língua portuguesa encontramos:

- (30 endereços)
- (6 endereços)
- (7 endereços)
- (12 endereços)

2.1 Apresentação do Corpus de Referência do Português Contemporâneo

O Corpus de Referência do Português Contemporâneo (CRPC) do Centro de Linguística da Universidade de Lisboa, iniciado em 1988, é um corpus linguístico, electrónico, que contém actualmente 334 milhões de palavras sendo constituído por amostragens de diversos tipos de texto de discurso escrito (literário, jornalístico, técnico, científico, didáctico, económico, jurídico, parlamentar, etc.) e de discurso oral (elocuições informais e formais).

Estas amostragens dizem respeito a variedades nacionais e regionais do português: estão incluídas, no *corpus*, amostragens do português europeu, português do Brasil, português dos cinco países africanos de língua oficial portuguesa (Angola, Cabo Verde, Guiné-Bissau, Moçambique, São Tomé e Príncipe), português de Macau, português de Timor-Leste e do português falado em Goa. Do ponto de vista cronológico, o *corpus* contém textos que vão desde a 2.^a metade do séc. XIX até 2006, sendo, na sua maior parte, posteriores a 1970.

2.2 Apresentação do corpus CETEMPúblico

O CETEMPúblico (Corpus de Extractos de Textos Electrónicos MCT/Público) é um corpus de aproximadamente 180 milhões de palavras em português europeu, criado pelo projecto Processamento computacional do português (projecto que deu origem à Linguateca) após a assinatura de um protocolo entre o Ministério da Ciência e da Tecnologia (MCT) português e o jornal PÚBLICO em Abril de 2000. O PÚBLICO é um jornal diário português de grande circulação, fundado em 1991, e o primeiro jornal português a disponibilizar uma edição electrónica na rede

A convincente maioria do texto do CETEMPúblico está em português europeu, embora haja alguns textos de autores brasileiros e africanos. O corpus inclui o texto de cerca de 2.600 edições do PÚBLICO, entre os anos de 1991 e 1998, num total de aproximadamente 180 milhões de palavras (versão 1.0). O CETEMPúblico 1.0 está dividido em 1.567.625 extractos, classificados por semestre e secção do jornal da qual provêm. Cada extracto está dividido em parágrafos e frases, e os títulos e os autores dos artigos estão assinalados.

2.3 Apresentação do corpus paralelo bi-direccional COMPARA

Este corpus tem como base textos em português e as suas traduções para inglês e textos em inglês e as suas traduções para português. Pode ser utilizado para pesquisar como foram traduzidas palavras e expressões de português para inglês e de inglês para português.

A pesquisa simples permite pesquisar palavras e expressões em todo o corpus; os resultados são apresentados em forma de concordâncias paralelas. A pesquisa avançada permite efectuar pesquisas mais sofisticadas, seleccionar textos específicos, escolher as partes do COMPARA que o utente pretende utilizar (se não quiser usar todos os textos do corpus) e seleccionar e obter diferentes tipos de resultados.

O COMPARA é um corpus extensível e tem como objectivo incorporar um número cada vez maior e mais representativo de traduções, que possam ser disponibilizadas à medida que vão sendo incorporadas no corpus. Neste momento, o COMPARA dispõe de autorizações para incluir excertos de 62 pares de textos de ficção contemporânea e não contemporânea, de autores e tradutores da África do Sul, Angola, Brasil, Estados Unidos, Moçambique, Portugal e Reino Unido, e muitos destes textos já estão disponíveis no corpus.

3 Procedimento de análise de série de concordâncias, exemplo da série de concordâncias extraídas do corpus Compara

A nossa condição de procura foi, para começar, o lema »tênis«, a seguir o lema »tênis«. Optámos por esta forma lexical à base das definições de dois dicionários monolíngues da língua portuguesa. O Dicionário Editora da Língua Portuguesa, que é um dos dicionários do português europeu, define a entrada “tênis» da maneira seguinte:

tênis: substantivo masculino 2 números, DESPORTO, jogo com uma bola de borracha revestida de camurça, impelida por uma raqueta, praticado em pavimento próprio, dividido ao meio por uma rede

O português europeu só conhece a forma ortográfica ténis e o significado descrito mais acima. O Dicionário Aurélio, que é um dos dicionários monolíngues mais importantes da variedade da língua portuguesa do português do Brasil, contém a entrada ténis mas explicada como lusitanismo. É por isso que nos dirige para a entrada »tênis«. Para esta entrada expõe a seguinte explicação:

Tênis: S. m. 2 n.

1. Jogo de origem inglesa, com raquetes e bola, em campo adrede preparado, dividido em duas partes por uma rede de malhas por cima da qual a bola deve passar.
2. Sapato de lona, couro, etc., us. na prática do ténis e de outros esportes, ou ainda com traje informal; sapato-tênis, basquete e (lus.) sapatilha.

O objeto da nossa investigação é o seguinte:

- verificar se a ortografia do português do Brasil também figura no português europeu e
- verificar se a forma ortográfica ténis tem o significado de »sapatilha« ou »sapato-tênis« também no português europeu.

A série de concordâncias do corpus COMPARA tem para o lema »tênis« 24 ocorrências.

A condição da procura: “tênis”:

- | | | |
|---------------|---|--|
| EBDL1T2(23): | Raramente vinha quando parecia ser mais lógico, como num jogo de golfe ou de tênis , mas podia vir logo <i>depois</i> do jogo, no bar do clube, ou enquanto dirigia de volta para casa, ou quando estava sentado no estúdio, ou deitado na cama. | It rarely happened when I might have expected it, like when I was playing golf or tennis, but it could happen just <i>after</i> a game, in the club-house bar, or while driving home, or when I was sitting perfectly still in my study, or lying in bed. |
| EBDL1T2(507): | Não tem nada que se compare com aquele cansaço dolorido e gostoso que a gente sente depois de um jogo de <i>squash</i> ou dezoito buracos no golfe ou cinco <i>sets</i> no tênis , a sensação reconfortante de esticar os braços e as pernas entre os lençóis quando se vai para a cama, sabendo que se deslizará sem esforço num sono profundo e longo. | I don't know anything like that glowing, aching tiredness you feel after a keen game of squash or eighteen holes of golf or five sets of tennis, the luxury of stretching out your limbs between the sheets when you go to bed, knowing you're just about to slide effortlessly into a long, deep sleep. |

- EBDL1T2(537): Mas ainda jogo um pouco de **tênis**, usando um tipo de joelheira que o mantém mais ou menos rígido. But I do play a bit of tennis still, wearing a kind of brace on the knee which keeps it more or less rigid.
- EBDL1T2(548): Fui ao clube outro dia com o meu grupo de deficientes físicos e eis que Sally apareceu, vindo direto do trabalho para uma aula de **tênis**. I was down at the Club the other day with my physically-challenged peer group when she turned up, having come straight from work for a spot of coaching.
- EBDL1T2(553): Sally fica doida quando percebe — com razão. Por isso quando acenou para mim casualmente através da rede, acenei de volta casualmente, com a cara de quem soubesse que ela combinara a aula de **tênis** para esta tarde. Sally gets pissed off when she twigs it — ununderstandably — so when she waaved casually to me through the netting, I waved back casually in case I was supposed to know that she had arranged to have coaching that afternoon.
- EBDL1T2(563): Quando percebi que era ela no clube de **tênis**, senti um certo orgulho de possessão de sua figura esbelta e seus cachos loiros. When I sussed it was her down at the tennis club, I felt a little glow of proprietorial pride in her lissome figure and bouncing golden locks.
- EBDL1T2(870): O Country Club, com seu campo de golfe de dezoito buracos, quadras de **tênis**, piscina coberta e ao ar livre e um spa, fica a apenas dez minutos de casa. The Country Club, with its eighteen-hole golf course, tennis courts, indoor and outdoor pools and spa, is just ten minutes away.
- EBDL1T2(1355): Imaginei minha vida sem sexo, sem **tênis**, sem um *show* de tevê. I contemplated a life without sex, without tennis without a TV show.
- EBDL1T2(1491): Sally e eu fomos lá para um jogo de **tênis** na quadra ao ar livre, depois de um almoço cedo. Sally and I went there after an early lunch to play tennis, outdoors.
- EBDL1T2(1824): De manhã, notei também uma pontada no cotovelo por causa do **tênis**, quando fazia a barba. This morning I noticed a touch of tennis elbow, too, when I was shaving.
- EBDL1T2(1825): Seria demais, não seria, se fizesse mais uma operação no joelho, só para descobrir que teria de parar de jogar **tênis** por causa do cotovelo? Be great, wouldn't it, if I had another operation on my knee, only to find I had to give up tennis because of the elbow?
- EBDL1T2(2082): Algum tempo depois, consegui produzir um jatinho de esperma por puro esforço físico, por isso agora sei que a parte mecânica está em ordem, mas meu pau ficou muito dolorido e, além disso, não foi nada bom para o meu reumatismo causado pelo **tênis**. Eventually I managed to produce a spurt of jism by sheer physical effort, so I know the plumbing is basically sound, but my cock is quite sore and it hasn't done my tennis elbow any good either.

- EBDL3T2(1151): Havia um programa de esportes de quatro horas de duração nas tardes de sábado, para o qual sentou-se em frente ao aparelho cheio de expectativa; mas parecia que havia uma espécie de conspiração para levar a população aos estádios de futebol ou ao supermercado ou a qualquer lugar, já que qualquer coisa era mais interessante que ver aquela rapidíssima sucessão de reportagens sobre esgrima feminina, campeonatos de natação locais, competições de pesca e torneios de **tênis** de mesa.
- There was a four-hour programme of sport on Saturday afternoons which he had settled down to watch expectantly, but it seemed to be some kind of conspiracy to drive the population out to the soccer stadiums or to the supermarkets or anywhere rather than watch ladies' archery, county swimming championships, a fishing contest and a table-tennis tournament all in breathtaking succession.
- PBCB2(94): À medida que fala, minha irmã espalha uma película de geléia grená na torrada, como que esmaltando a torrada, depois analisa, desiste do grená e arremata com geléia cor de laranja; vai morder, muda de idéia, toma um gole de chá e se admira de como uma pessoa pode envelhecer da noite para o dia, pois quando papai morreu a gente pensou que mamãe fosse baquear, qual nada, continuou a mil, ia ao teatro, jantava fora, andava a cavalo no sítio, como ela adorava o sítio, tomava seu uisquinho, jogava **tênis**, puxa, pensar que até o ano passado mamãe ainda jogava tênis.
- As she talks, my sister spreads a film of purple jam over her toast, as if she is enamelling the toast, then she takes stock, puts aside the purple jam and finishes off with the orange; she goes to take a bite, changes her mind, has a sip of tea and marvels at how people can grow old overnight, for when Father died people thought Mummy would go downhill, not a bit of it, she gathered steam, went to the theatre, dined out, enjoyed horse-riding on the farm, simply adored the farm, drank her drop of whisky, played tennis, drat it, to think that only last year Mummy was still playing tennis.
- PBMR1(610): Na altura do Rio (assinalado no mapa com uma cruzinha ou uma pequena maçã) atirei uma bola de **tênis** na direção do sr. Franco.
- Near Rio (shown on the map by a small cross or apple) I threw a tennis ball in Mr. Franco's direction.
- PBPC2(773): Petrus pediu que eu tirasse os **tênis** e a camisa.
- Petrus told me to take off my sneakers and my shirt.
- PBPM1(542): Fechava os olhos e podia vê-lo com suas roupas de **tênis**, meias esticadas até as canelas, a raquete no ombro.
- I could close my eyes and see him in his tennis outfit, socks up to his knees, the racket on his shoulder.
- PBPM2(391): O porquinho brincando com o meu **tênis**.
- The pig played with my sneaker.
- PBPM2(544): Um dia, voltei para casa e encontrei meu **tênis** branco todo estraçalhado.
- One day I came home and found my white sneakers all chewed up.
- PBPM2(555): O que aconteceu com este **tênis**?
- What happened to the sneakers?

- PBPM2(980): Esse **tênis** Redley, ele me deu, o meu amor. He gave me these sneakers, my true love.
- PBRF1(163): Meu pai era um homem bonito com muitas namoradas, jogava **tênis**, nada, nunca pegara uma gripe — até ter um derrame cerebral. My father was a handsome man with many girlfriends, a man who played tennis, swam and was never ill, even with a cold — until his cerebral haemorrhage.
- PBRF2(992): Eu havia saído para jogar **tênis** no Country, à tarde, mas começou a chover e voltei para casa e lá estava Val deitado na cama fazendo coisas com um amigo nosso. I was out playing tennis at the country club, but it started to rain, so I came back home, and there was Val in bed messing around with a guy we knew.

A série de concordâncias do corpus Compara para o lema »tênis« tem 30 ocorrências.

A condição de procura: »tênis«:

- EBDL1T1(23): Raramente aparecia quando seria de esperar, como, por exemplo, quando estava a jogar golfe ou **tênis**, mas podia senti-la imediatamente *a seguir* a uma partida, enquanto estava no bar do clube ou ia a guiar para casa, ou então quando estava sentado no escritório, perfeitamente imóvel, ou deitado na cama em casa. It rarely happened when I might have expected it, like when I was playing golf or tennis, but it could happen just *after* a game, in the club-house bar, or while driving home, or when I was sitting perfectly still in my study, or lying in bed.
- EBDL1T1(507): Não conheço nada que se equipare àquele cansaço afogado e doloroso que sentimos a seguir a um bom jogo de *squash*, a 18 buracos de golfe ou a 5 *sets* de **tênis**; nada igual à luxúria com que esticamos as pernas e os braços por entre os lençóis quando nos deitamos, sabendo que daí a pouco deslizaremos sem qualquer esforço para dentro de um sono longo e profundo. I don't know anything like that glowing, aching tiredness you feel after a keen game of squash or eighteen holes of golf or five sets of tennis, the luxury of stretching out your limbs between the sheets when you go to bed, knowing you're just about to slide effortlessly into a long, deep sleep.
- EBDL1T1(537): Mas continuo a jogar um pouco de **tênis**, com uma espécie de liga no joelho, que o mantém mais ou menos rígido. But I do play a bit of tennis still, wearing a kind of brace on the knee which keeps it more or less rigid.
- EBDL1T1(539): No clube há *courts* de **tênis** cobertos, mas de qualquer forma é possível jogar em campos descobertos quase todo o ano, com os invernos suaves que ultimamente temos tido — parece ser um dos poucos efeitos benéficos do aquecimento da Terra. They have indoor courts at the Club, and anyway you can play outdoors nearly all the year round with these mild winters we've been having — it seems to be one of the few beneficial effects of global warming.

- EBDL1T1(548): Noutro dia estava no clube com o meu grupo de jogadores- que-arriscam o físico quando ela apareceu, vinda directamente do emprego para uma lição de **ténis**.
I was down at the Club the other day with my physically-challenged peer group when she turned up, having come straight from work for a spot of coaching.
- EBDL1T1(563):) Quando percebi que era ela, lá no clube de **ténis**, senti um pequeno vislumbre de orgulho por aquela figura ágil e pelos seus caracóis louros ondulantes.
When I sussed it was her down at the tennis club, I felt a little glow of proprietorial pride in her lissome figure and bouncing golden locks.
- EBDL1T1(870): O clube, com um campo de golfe de 18 buracos, *courts* de **ténis**, piscinas cobertas e descobertas e hidromassagem, fica apenas a dez minutos de casa.
The Country Club, with its eighteen-hole golf course, tennis courts, indoor and outdoor pools and spa, is just ten minutes away.
- EBDL1T1(1355): Vi-me perante uma vida sem sexo, sem **ténis** e sem o programa de televisão.
I contemplated a life without sex, without tennis without a TV show.
- EBDL1T1(1491): Eu e a Sally almoçámos cedo e depois fomos lá para jogar **ténis** ao ar livre.
Sally and I went there after an early lunch to play tennis, outdoors.
- EBDL1T1(1824): Hoje de manhã, quando estava a fazer a barba, tive também uma cáibra no cotovelo, daquelas que são provocadas pelo **ténis**.
This morning I noticed a touch of tennis elbow, too, when I was shaving.
- EBDL1T1(1825): Não era engraçado ser outra vez operado ao joelho e descobrir depois que tinha de deixar de jogar **ténis** por causa do cotovelo?
Be great, wouldn't it, if I had another operation on my knee, only to find I had to give up tennis because of the elbow?
- EBDL2(1186): Durante esta discussão, as cabeças dos outros homens presentes têm estado a virar-se de um lado para outro, como espectadores de uma partida de **ténis**.
The heads of the other men present have been swivelling from side to side, like spectators at a tennis match, during this argument.
- EBDL3T1(1151): Havia um programa desportivo de quatro horas ao sábado à tarde, que esperava ansiosamente ver, mas que parecia uma espécie de conspiração para levar a população até aos estádios, supermercados ou onde quer que fosse, de preferência senhoras no tiro ao arco, campeonatos regionais de natação, concursos de pesca e torneios de **ténis** de mesa, tudo numa sucessão alucinante.
There was a four-hour programme of sport on Saturday afternoons which he had settled down to watch expectantly, but it seemed to be some kind of conspiracy to drive the population out to the soccer stadiums or to the supermarkets or anywhere rather than watch ladies' archery, county swimming championships, a fishing contest and a table-tennis tournament all in breathtaking succession.

- EBDL4(913): Inscreveu-se num clube de **ténis**, mas as raparigas acharam-no afectado e pedante. He joined a tennis club, but the girls there found him stiff and priggish.
- EBDL4(914): Também não era muito bom na dança nem no **ténis**. He was not very good, either, at dancing or tennis.
- EBDL6(384): Pelas janelas abertas entram os ruídos de actividades estivais, o poc-poc das bolas de **ténis** saltitando de um lado para o outro nos *courts* vizinhos, o grito ocasional «Com' équ'estáisso?» que se ouve nos campos de jogos, e os acordes de violas e harmónicas de um serão musical, improvisado ao ar livre junto ao lago artificial. Through its open windows carry the sounds of summer pursuits, the *pock pock* of tennis balls bandied to and fro on the nearby courts, the occasional cry of *Howzat* ? from the playing fields, and the sounds of guitars and harmonicas from an impromptu al fresco singalong down by the artificial lake.
- EBDL6(477): Morris muda de roupa, vestindo os calções de cetim vermelhos, a *sweat-shirt*, com o emblema de Euphoric State, calça os **ténis** *Adidas* e, ao sair, enfia a carta no marco do correio que está no vestíbulo da *villa*. Morris changes into his red silk running shorts and Euphoric State sweat-shirt and Adidas training shoes, and drops his letter into the mail box in the hall on his way out of the villa.
- EBIM1(141): Era aquilo o juramento do campo de **ténis**? Was this the Tennis Court Oath?
- EBIM2(191): No mundo reinava a barafunda habitual: os peixes estavam a mudar de sexo, o **ténis** de mesa inglês estava a perder o rumo e, na Holanda, certos tipos repugnantes, formados em Medicina, estavam a oferecer serviço legal para eliminar os parentes idosos incómodos. The world was its usual mess: fish were changing sex, British table tennis had lost its way, and in Holland some unsavoury types with medical degrees were offering a legal service to eliminate your inconvenient elderly parent.
- EBJB3(636): Não consegui deixar uma reprodução exacta da cara dela no depósito de bagagens da memória; mas *creio* que vestia uma camisa dum tom entre verde casto e lilás louco, sobre uns jeans 501 cinzentos desbotados, meias verdes e **ténis** de corrida ridículos, inestéticos. I failed to deposit an accurate simulacrum of her visage and demeanour with the left-luggage clerk of memory; but I *think* she was in a shirt of a hue between sage and lovage, atop grey stone-washed 501s, green socks and a ridiculously unaesthetic pair of trainers.
- EBJB3(647): Antes de saírem, não pude deixar de me abeirar da Gillian e de lhe dar o precioso conselho de que usar jeans 501 com **ténis** de corrida era francamente *un désastre* que me espantava como tinha conseguido vir a andar pela rua fora em pleno dia até o meu apartamento, sem ir parar à força. Before they left I could not forbear to draw Gillian closer and impart to her the glittering counsel that wearing 501s with trainers was frankly *un désastre* and that I was amazed she had walked the streets to my apartment in broad daylight and escaped pillory.

EJB3(721):	Desagrada-lhe que use ténis de corrida com os jeans?»	Do you mind if I wear trainers with 501s?»
EJBT1(564):	Tinha umas calças de malha roxas, uma camisola preta, demasiado curta, e ténis turquesa,	She wore purple leggings and a black jersey — not long enough – and turquoise trainers.
EJBT3(791):	Eram como os ténis , que nunca eram engraxados.	It was like these sports-shoe things that'd never take a polish.
EJBT3(1694):	À saída viu que na sala de espera estava um homem ainda novo, com <i>jeans</i> e ténis , um rosto como uma caveira e uma cabeça quase rapada.	On the way out, she saw that there was a young man in the waiting room, in jeans and trainers, with a face like a skull and an almost shaven head.
EBKI1(1582):	Depois seguiu na direcção da cidade, passando pelos campos de ténis e pelas zonas comerciais.	Then he went towards the town, past the tennis courts and through the shopping areas.
ESNG2(140):	Num campo de ténis abandonado, a reluzir de lustrosas ervas daninhas, um pássaro plangente pressagiava chuva.	From an abandoned tennis court brilliant with glossy weeds a mournful bird presaged rain.
ESNG2(506):	Mas nessa tarde perguntou a William Donaldson se ele lhe dava a oportunidade de o vencer numa partida de ténis : dizia-se por piada, quando ela estava com os Donaldson, que, embora ele jogasse assiduamente num clube desportivo de homens de negócios para manter a forma, nunca ganhara uma partida a não ser contra ela.	But that afternoon she asked William Donaldson whether he would give her a chance to beat him at a game of tennis — it had been a joke, when she stayed with the Donaldsons, that although he played assiduously at some businessmen's sports club to keep fit, he never won a set with anyone but her.
ESNG2(1087):	Deambulava por aquele jardim abandonado — a deescendência da velha Lolita apanhava deuses hotentotes na relva que invade o campo de ténis — e sabia que devia ter-lhe desejado a morte; que exultar e sofrer eram para mim a mesma coisa.	I prowled about that abandoned garden, old Lolita's offspring caught Hot-tentot Gods in the grass that had taken over the tennis court, and I knew I must have wished him to die; that to exult and to sorrow were the same thing for me.
ESNG2(1149):	As ervas daninhas cresciam luxuriantes no sítio onde eu tropeçava em tranças de papel de jornal sujo de merda humana, garrafas e farrapos arremessados para o meio dos arbustos aromáticos onde se costumavam perder bolas de ténis .	Weeds broke rank where I tramped over twists of newspaper smeared with human shit, bottles and rags cast among the scented shrubs where tennis balls used to be lost.

3.1 As constatações ao consultar as séries de concordâncias para os lemas »ténis« e »tênis« segundo o objeto da nossa investigação: se a forma ortográfica do português do Brasil também figura no português europeu.

A comparação de duas séries de concordâncias para »ténis« e »tênis« demonstra que as concordâncias são muito parecidas, contudo, não iguais. Há mais exemplos deste tipo nas séries de concordâncias. Vejamos um exemplo escolhido ao acaso:

Imaginei minha vida sem sexo, sem **tênis**, sem I contemplated a life without sex, without tennis without a TV show.
um *show* de tevê.

EBDL1T2 David Lodge

1995 *Therapy* London: Secker & Warburg, pp 3-97. Copyright © 1995 David Lodge.

Uso autorizado por David Lodge.

ORIGINAL: Inglês, Reino Unido; 43742 unidades, 37664 palavras, 6874 tipos, **2149** unidades de alinhamento.

1997 *Terapia* tradução de Lídia Cavalcante-Luther. São Paulo: Scipione, pp 11-115. Copyright © 1997 Scipione. Uso autorizado por Editora Scipione.

TRADUÇÃO: Português, Brasil; 45192 unidades, 39103 palavras, 7534 tipos.

EBDL1T1 David Lodge

Vi-me perante uma vida sem sexo, sem **ténis** e I contemplated a life without sex, without tennis without TV show.
sem o programa de televisão.

1995 *Therapy* London: Secker & Warburg, pp 3-97. Copyright © 1995 David Lodge.

Uso autorizado por David Lodge.

ORIGINAL: Inglês, Reino Unido; 43742 unidades, 37664 palavras, 6874 tipos, **2149** unidades de alinhamento.

1995 *Terapia* Tradução de Maria do Carmo Figueira. Lisboa: Gradiva, pp 11-88. Copyright © 1995 Gradiva Publicações, Lda. Uso autorizado por Gradiva Publicações, Lda.

TRADUÇÃO: Português, Portugal; 45165 unidades, 38991 palavras, 7919 tipos.

Depois de termos verificado os dados que demonstram a quem pertencem os textos, constatámos que se trata da tradução da mesma obra literária, só que a tradução que contém a forma ortográfica »tênis« é uma tradução para a variedade do português do Brasil. O texto onde está presente a forma ortográfica »ténis« é a tradução para o português europeu.

Ao seguir verificamos as ocorrências das concordâncias para o lema »ténis« com a pesquisa avançada. Este tipo de pesquisa oferece a possibilidade da procura das concordâncias sob diversas condições restritivas. A nossa condição restritiva ou seja limitativa na procura que pretendemos desta vez foi a pesquisa dentro dos textos da variedade do português do Brasil. Para o lema »ténis« não encontramos nenhuma ocorrência.

Também na procura »tênis« impusemos uma limitação. Desta vez pretendemos o tipo da procura das concordâncias restringida, ou seja a procura pelas partes do corpus que constam dos textos escritos no português europeu. Tampouco houve ocorrências.

O resultado do tipo de pesquisa avançada demonstra que a forma ortográfica »ténis« no COMPARA está ligada ao português europeu, a forma ortográfica »tênis« exclusivamente à variedade brasileira desta língua.

3.2 As constatações ao consultar as séries de concordâncias para os lemas »tênis« e »tênis« segundo a nossa investigação: se a palavra ténis tem o significado de »sapatilha« ou »sapato-tênis« também no português europeu.

Depois de termos efectuado o procedimento da análise do corpus podemos concluir que a forma lexical ou seja a forma ortográfica »tênis« definitivamente tem o significado de »sapato – tênis« ou »sapatilha«, embora o dicionário monolíngue do português europeu não refira este significado para esta palavra.

O primeiro exemplo encontramos sob a nota de EBDL6 (477) que revela a tradução da seguinte obra literária:

EBDL6 David Lodge

1984 *Small World*. Londres: Penguin (1985), pp. 231-333. Copyright © David Lodge 1984. Uso autorizado por David Lodge.

ORIGINAL: Inglês, Reino Unido; 48236 unidades, 39732 palavras, 7462 tipos, **2796** unidades de alinhamento.

1996 *O Mundo é Pequeno*. tradução de Lucinda Santos Silva. Porto: Edições Asa, pp. 251-355. Direitos portugueses reservados, Edições Asa. Uso autorizado por Edições Asa.

TRADUÇÃO: Português, Portugal; 49667 unidades, 41121 palavras, 8915 tipos.

É o caso da tradução para o português europeu, a obra foi publicada em Portugal.

O segundo exemplo encontramos sob a nota de EBJB3 (636):

EBJB3 Julian Barnes

1991 *Talking it Over* Versão baseada no excerto digitalizado pelo ENPC a partir de Londres: Jonhathan Cape, pp 1-46. Copyright © 1991 Julian Barnes. Uso autorizado por Julian Barnes.

ORIGINAL: Inglês, Reino Unido; 15431 unidades, 13294 palavras, 3181 tipos, **942** unidades de alinhamento.

1994 *Amor e Co.* tradução de Helena Cardoso. Versão baseada no excerto digitalizado pelo ENPC a partir de Lisboa: Quetzal, pp 9-49. Copyright © 1994 Editora Quetzal. Uso autorizado por Quetzal.

TRADUÇÃO: Português, Portugal; 15278 unidades, 12869 palavras, 3749 tipos.

Também aqui se trata da tradução para o português europeu, a obra foi publicada em Portugal.

O seguinte exemplo tem a nota de EBJT1 (564)

EBJT1 Joanna Trollope

1996 *Next of kin* Texto digitalizado a partir da edição London: Black Swan, 1997, pp 7-87. Copyright © 1996 Joanna Trollope. Uso autorizado por Joanna Trollope.

ORIGINAL: Inglês, Reino Unido; 32483 unidades, 28105 palavras, 4586 tipos, **2078** unidades de alinhamento.

1998 *Parentes próximos* tradução de Ana Falcão Bastos. Lisboa: , pp 7-77. Copyright © 1998 Gradiva Publicações, Lda. Uso autorizado por Gradiva Publicações, Lda.

TRADUÇÃO: Português, Portugal; 32611 unidades, 27215 palavras, 5709 tipos.

Também esta tradução foi publicada em Portugal no português europeu.

À base das conclusões que provêm da investigação das concordâncias para os lemas »tênis« e »ténis« listadas seguindo o tipo da pesquisa simples, podríamos constatar que a palavra »ténis« tem o significado de »sapatilha« também no português europeu. O facto é que o Dicionário Porto Editora ainda não regista este significado.

Convinha verificar este significado também no Corpus de Referência do Português Contemporâneo. Dado que este inclui os textos de todas as variantes do português, seria melhor verificar o significado da forma lexical »ténis« no corpus CETEMPúblico, que é um dos corpora de textos em português europeu.

Realizamos a procura da série de concordâncias para o lema »ténis« no corpus CETEMPúblico.

Resultados da procura

Procura: “**ténis**”. Pedido de uma concordância em contexto Corpus: CETEMPúblico 1.7 5417 ocorrências.

Consultámos só as primeiras trezentas ocorrências e constatámos que entre elas encontramos o lema »ténis« também com o significado de »sapatilha«, embora o corpus CETEMPúblico contenha sobretudo os textos escritos no português europeu. Referimos alguns exemplos:

Ext 21447 (nd, 95a): «Todas as crianças deviam aprender a atar sozinhas os cordões dos **ténis** porque isso, que parece uma simples coisinha, desenvolve aptidões importantes.»

Ext 31944 (nd, 95a): Está muito bronzeado e a roupa é, no mínimo, informal: calções, um pólo azul escuro e, a rematar, sapatos de **ténis** .

Ext 35005 (soc, 93b): Tem uma câmara de vídeo Sony a tiracolo, usa Lewi ' s 501 e **ténis** Reebok .

Ext 38054 (clt, 92b): A completar o conjunto, **ténis** e meias de lã .

Ext 48616 (des, 98a): Neste jogo, o atleta calçou os seus primeiros **ténis** da NBA e declarou que aqueles **ténis** lhe tinham provocado «enormes bolhas nos pés», o preço a pagar por «regressar a 1984» .

Ext 48688 (soc, 98b): De calções, t-shirt e **ténis**, os dois amigos tinham um «look» todo desportista.

No corpus CETEMPúblico procurámos também a série de concordâncias para o lema “**ténis**”, ou seja para a forma ortográfica brasileira desta palavra. Dado que este corpus é um corpus dos textos escritos no português europeu, achámos uma ocorrência só.

Resultados da procura

Procura: “**ténis**”, Pedido de uma concordância em contexto, Corpus: CETEMPúblico 1.7

Uma ocorrência.

Ext 1059115 (des, 98a): As tendas VIP dos grandes acontecimentos desportivos do ténis, do golfe — ««tênis» e «gólfe», para quem quiser desde já entrar no espírito — obedecem a esta verdade absoluta e ponto final.

4 Conclusão

A conclusão da nossa pesquisa dos lemas “tênis” e “ténis” tanto no corpus COMPARA como no corpus CETEMPúblico é a seguinte: à base da investigação da série de concordâncias em contexto para o lema “ténis” podemos constatar que esta palavra tem no português europeu salvo o significado »tênis- jogo de desporto« também o significado de »sapatilha«, só que este significado ainda não está presente na explicação do termo no dicionário monolíngue do português europeu Dicionário Editora da Língua Portuguesa. Também constatámos que o lema “tênis” está presente só nos textos escritos na variedade brasileira do português. O lema “ténis” podemos encontrar só naqueles textos do corpus que são escritos no português europeu. A análise do corpus confirma a nossa tese e a descrição linguística geralmente aceite, que a forma ortográfica “tênis” pertence ao português do Brasil ao mesmo tempo que a forma ortográfica “ténis” pertence ao português europeu, contudo, não confirma a definição lexicográfica exposta no Dicionário Editora da Língua Portuguesa que diz que a entrada “ténis” tem um significado só, o significado do jogo desportivo.

BIBLIOGRAFIA

Gorjanc, Vojko (2005): *Uvod v korpusno jezikoslovje*. Domžale: Izolit

Erjavec, Tomaž (1996/97): »Računalniške zbirke besedil.« Em: *Jezik in slovstvo* 2/3, 81-95

Bacelar do Nascimento, M. F. (2003): »« Em: Mendes, A. et alii (orgs.) *Actas do XVIII Encontro da Associação Portuguesa de Linguística*, Lisboa: Associação Portuguesa de Linguística e Edições Colibri.

<http://www.linguateca.pt/>

<http://www.linguateca.pt/ACDC/>

<http://www.dicionarios-online.com/>

http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_crpc.php

<http://www.linguateca.pt/CETEMPUBLICO/>

<http://www.linguateca.pt/COMPARA>

PREDSTAVITEV JEZIKOVNIH VIROV
ZA PORTUGALSKI JEZIK
IN POSTOPKI KORPUSNE ANALIZE
PORTUGALSKEGA BESEDIŠČA

Avtorica članka na splošno ugotavlja, da so na Portugalskem temelji za gradnjo korpusov in drugih virov o jeziku trdno postavljeni in da so v teku zanimivi projekti. Obstoječi korpusi vključujejo vse variante portugalščine, obstoječi programi pa so dovolj prijazni, da omogočajo preprostejše postopke korpusne analize vsakemu uporabniku. Avtorica uporabi nekaj postopkov korpusne analize in na podlagi pregleda niza konkordanc za različnico "tênis" zapiše, da ima ta v evropski portugalščini poleg pomena tenis »športna igra« tudi pomen »športni copat«, ki pa ga enojezični slovar evropske portugalščine Dicionário Editora da Língua Portuguesa ne navaja in tako, kakor ugotavlja avtorica, ne odslíkava realnega stanja v jeziku. Avtorica prav tako ugotavlja, da se različnica "tênis" v korpusu pojavlja le v besedilih, pisanih v brazilski portugalščini, različnico "tênis" pa je v korpusu zaslediti le v tistih besedilih, ki so napisana v evropski portugalščini. Korpusna analiza tako podpre njeno tezo in uveljavljeni jezikovni opis, da zapis besede "tênis" pripada brazilski varianti portugalskega jezika, zapis besede "tênis" pa evropski portugalščini, ne potrdi pa slovarskega opisa slovarja Dicionário Editora da Língua Portuguesa, da ima geslo "tênis" le pomen tenis »športna igra«.