

Convergência e divergência no léxico do Português Europeu e do Português Brasileiro: resultados do estudo sobre termos de futebol e de moda

Augusto Soares da Silva
Universidade Católica Portuguesa – Braga

1. Introdução

O objectivo principal deste estudo é apresentar os resultados do projecto de investigação *Convergência e Divergência no Léxico do Português* relativamente aos domínios do futebol e da moda.¹ Observando a variação lexical onomasiológica de termos destes dois domínios, ao longo dos últimos 50 anos, vamos medir a distância entre o Português Europeu e o Português Brasileiro e, assim, saber se as duas variedades se encontram num processo de convergência ou divergência lexical. No plano teórico e metodológico, e na sequência de textos programáticos (Silva, 2004a, b, 2005), pretendemos mostrar as vantagens da perspectiva *cognitiva* (da Linguística Cognitiva, de R. Langacker, G. Lakoff e L. Talmy) e de métodos *quantitativos* no estudo das variedades *lectais* – neste caso, variedades nacionais de uma língua transcontinental.

Num primeiro momento, daremos conta das questões da presente investigação e das hipóteses existentes. A seguir, exporemos os elementos essenciais do enquadramento teórico e metodológico deste estudo de sociolinguística cognitiva e quantitativa do Português. Finalmente, apresentaremos os dados e resultados relativamente aos termos de futebol e de moda/vestuário.

2. Questões e hipóteses

A questão principal da presente investigação é saber se o Português Europeu e o Português Brasileiro (doravante, PE e PB) se encontram num processo de convergência ou de divergência a nível do léxico. Esta questão implica obviamente uma análise diacrónica. O período de análise abrange as últimas 5 décadas. Dentro desta questão diacrónica, uma questão *interna* é saber que influência têm factores como os estrangeirismos na tendência evolutiva geral.

¹ O projecto é financiado pela Fundação para a Ciência e a Tecnologia (Ref^o POCTI/LIN/48575/2002), no âmbito do POCTI do fundo comunitário europeu FEDER, e tem a duração de dois anos (Janeiro 2004 – Janeiro 2006). Integram a equipa de investigação, para além do autor deste texto, José João Dias Almeida, Alberto Manuel Simões, Ana Margarida Abrantes, Marlene Danaia Duarte, José Luiz de Lucca e Ana Margarida Nunes.

Secundariamente, queremos responder à questão sincrónica da estratificação lexical actual das duas variedades nacionais. Mais especificamente, pretendemos saber se a distância lexical entre registo padrão e registos intermédios é maior no PB ou no PE.²

Como hipóteses sobre as relações entre PE e PB, admite-se (i) uma maior distância entre registos formais e informais no PB do que no PE; (ii) uma influência crescente do PB sobre o PE (e as variedades africanas), que costuma ser justificada pelas tão populares telenovelas brasileiras, pelo sempre popular e apaixonante futebol brasileiro e pela emigração brasileira (Caetano, 2002), e (iii), embora não se encontrem hipóteses claras nem, muito menos, bem fundamentadas na respectiva literatura, a ideia generalizada de uma fragmentação progressiva e inelutável e, portanto, a hipótese da divergência.

Num jornal de referência, o semanário *Expresso*, o colunista Francisco Belard escreveu: “Continua a haver uma só língua portuguesa, mas nos últimos 50 anos, digamos, têm-se acentuado as diferenças na sintaxe e no léxico – sobretudo no vocabulário corrente” (Belard, 2001). E o director do mesmo influente semanário, José António Saraiva, num editorial de 24 de Maio de 1986, aquando do Acordo Ortográfico entre os sete países lusófonos, afirmava: “sendo sabido que o português falado em Portugal e no Brasil tenderá a divergir cada vez mais, caminhando-se no sentido da existência de duas línguas de facto diferentes” (Saraiva, 1986). A mesma previsão da fragmentação da língua portuguesa é claramente expressa pelo linguista Ivo Castro, escudando-se no que aconteceu ao latim do declinante Império Romano: “o modelo latino parece legitimar a previsão de que línguas como o inglês, ou o português, faladas em espaços muito vastos, habitados por povos que nada têm a ver historicamente com o berço da língua, se acham destinadas a uma progressiva fragmentação interna” (Castro, 1986: 45).

3. Enquadramento teórico e metodológico: sociolexicologia cognitiva e quantitativa

A presente investigação apoia-se na concepção geral e nos métodos quantitativos da investigação sociolexicológica desenvolvida por Dirk Geeraerts e sua equipa para o Neerlandês (Geeraerts, Grondelaers & Speelman, 1999).

O objecto de investigação é a variação lexical *onomasiológica*, isto é, a existência de diferentes termos para nomear determinado conceito ou referente.³ Mais especificamente, a variação onomasiológica *pragmática* – aquela que incide no uso e não na estrutura e tem a ver com as escolhas actuais de determinado nome entre vários possíveis como designação de determinado conceito ou referente. Ainda mais especificamente, a variação onomasiológica *pragmática formal* – aquela que envolve sinónimos referenciais, como *avançado*, *atacante* e *dianteiro* ou *casaco* e *blazer*, e não a variação

² Uma terceira questão, para a qual ainda não dispomos dos resultados de um inquérito, é saber como é que os falantes dos dois países conceptualizam a distância entre as duas variedades nacionais.

³ Sobre o mapa conceptual da variação lexical, ver Silva (2005).

onomasiológica conceptual, a qual envolve diferenças conceptuais, como *atacante* e *joagador* ou *casaco* e *blusão*. O que define as diferenças entre os sinónimos referenciais é a variação *contextual*, designadamente a variação geográfica, social, estilística, discursiva e diacrónica. É precisamente a variação onomasiológica formal e, dela fazendo parte integrante, a variação contextual o objecto mais específico da *sociolexicologia* (Silva, 2004a, 2005) e da presente investigação.

Porquê a variação onomasiológica formal como perspectiva preferencial para o estudo da convergência e divergência entre PE e PB? Metodologicamente, tem a vantagem de se poder estabelecer a distinção entre a frequência dos termos e a frequência dos conceitos. Linguisticamente, os sinónimos referenciais evidenciam tipos de diferenças sociolinguísticas. E são as diferenças sociolinguísticas – regionais, sociais, estilísticas, pragmático-discursivas e históricas – que definem a própria existência e a competição de variedades de uma língua.

Como medir a convergência ou divergência lexical entre duas variedades de uma língua? Para além de uma metodologia *baseada-no-uso* e, portanto, num *corpus* representativo, que permita a observação das escolhas onomasiológicas efectivamente realizadas pelos falantes entre alternativas disponíveis, o método específico é o da *uniformidade linguística baseada em perfis* (Geeraerts, 2001; Speelman, Grondelaers & Geeraerts, 2003). Este método assenta em duas noções essenciais: *perfil onomasiológico* e *uniformidade*.

Por *perfil onomasiológico* de um conceito/referente numa determinada variedade lectal entende-se o conjunto de sinónimos usados para designar esse conceito/referente nessa variedade, diferenciados pela sua frequência relativa. A *uniformidade* é a medida da correspondência entre dois conjuntos de dados, definidos em termos de perfis onomasiológicos. Por exemplo, a uniformidade de um conceito entre duas amostras, em que uma contém 6 ocorrências do termo A e 4 do termo B e a outra 3 ocorrências do termo A e 7 do termo B, resulta do número de pares comuns de nomeação desse conceito (7 pares), sendo portanto de 70%. Este mesmo resultado obtém-se somando as frequências relativas mais pequenas de cada termo alternativo: 30% do termo A (3 usos, provenientes da segunda amostra) e 40% do termo B (4 usos, provenientes da primeira amostra).

Em termos técnicos, o cálculo de uniformidade de um conceito é formulado assim:

$$U_z(Y_1, Y_2) = \sum_{i=1}^n \min(F_{z,y_1}(x_i), F_{z,y_2}(x_i))$$

A uniformidade U para um conceito Z entre duas amostras Y_1 e Y_2 equivale à soma \sum dos mínimos das frequências relativas F do termo x nos perfis onomasiológicos de Z em Y_1 e Y_2 . O símbolo x_i representa os diferentes termos x_1 a x_n usados nas amostras Y para designar o conceito Z .

Quando estão em causa vários conceitos, a uniformidade média é calculada em termos de *média ponderada*, através da seguinte fórmula:

$$U'(Y_1, Y_2) = \sum_{i=1}^n U_z(Y_1, Y_2) \cdot G_z$$

A uniformidade U' para um conjunto de conceitos Z entre duas amostras Y_1 e Y_2 equivale à soma dos valores- U dos Z s ponderados pela frequência relativa G de Z dentro do conjunto total de Z s. Obtém-se o coeficiente de ponderação G_{Z_i} ou frequência relativa de um conceito Z , dividindo a soma das ocorrências de Z em Y_1 e Y_2 pelo número total de ocorrências para os conceitos Z_1 a Z_n .

Resumindo, o cálculo da conformidade onomasiológica entre duas variantes de uma língua faz-se medindo o grau de uniformidade entre perfis onomasiológicos dessas variantes. Diacronicamente, convergência e divergência traduzem-se em aumento e diminuição da uniformidade, respectivamente. Sincronicamente, a distância maior ou menor entre registos formais e informais manifesta-se, respectivamente, em uniformidade menor e maior.

A orientação teórica é, como já foi referido, a da Linguística Cognitiva. Porquê ou quais as vantagens da perspectiva cognitiva nesta investigação sociolexicológica, ou numa outra qualquer? Primeiro, a Linguística Cognitiva caracteriza-se por assumir uma perspectiva *recontextualizante* (reintegrando no estudo da linguagem as diferentes formas de *contexto*, excluídas pelos modelos gramaticais autonomistas, particularmente o generativista), *experientialista*, orientada para o *significado* e *baseada-no-uso* (Silva 2004c). Resulta daqui a inevitabilidade de uma sociolexicologia cognitiva, em particular, e de uma sociolinguística cognitiva, em geral. Segundo, a flexibilidade e a variação são temas centrais neste paradigma. Finalmente, a Linguística Cognitiva, provavelmente mais do que qualquer outra abordagem contemporânea da linguagem, reconhece explicitamente, não só que a capacidade para a linguagem se fundamenta em capacidades cognitivas gerais, como também que todas estas capacidades são cultural e socialmente situadas e definidas. Mais concretamente, o estudo do papel da conceptualização da realidade social na constituição dos factos sociolinguísticos é uma das principais tarefas de uma linha de investigação conhecida como teoria dos *modelos culturais*. Ora, a sociologia da linguagem envolve justamente modelos culturais sobre variedades e variantes de uma língua, comportamentos e atitudes para com essa variação.

4. Materiais e o *corpus* CONDIVport

A base empírica da presente investigação consiste em largos milhares de observações do uso de termos alternativos (sinónimos referenciais) para nomear 43 conceitos nominais (substantivos) dos campos lexicais do futebol e da moda/vestuário. A base de dados resultante compreende pois as frequências de cada um dos termos dos 43 *perfis onomasiológicos* em textos portugueses e brasileiros. Foram escolhidos os domínios do futebol e da moda/vestuário por serem populares e permeáveis à influência de línguas estrangeiras, com a diferença da emotividade e presumível influência brasileira do primeiro.

Os materiais são extraídos de três fontes: (i) jornais e revistas de desporto e moda dos primeiros anos das décadas de 50, 70 e 90-2000; (ii) linguagem da Internet (conversação *on-line* de IRC ou *chats*); e (iii) etiquetas de lojas de vestuário de diferentes cidades tanto de Portugal como do Brasil. Os materiais de (i) permitirão

responder à questão da convergência/divergência e os materiais de (ii) e (iii), à questão da estratificação. Todos os materiais de (i) e (iii) foram extraídos manualmente.

O sub-corpus de futebol compreende textos dos jornais portugueses *A Bola*, *Record*, *Mundo Desportivo* e *O Jogo* e dos jornais brasileiros *Jornal dos Sports* (Rio de Janeiro), *Gazeta Esportiva* (São Paulo), *Estado de São Paulo* e *Lance*, com a extensão de 2,7 milhões de palavras, e textos de *chats* de canais associados a clubes de futebol portugueses e brasileiros, recolhidos entre Abril e Outubro de 2005, com a extensão de 15 milhões de palavras.

O sub-corpus da moda compreende textos de 28 revistas de moda, com a extensão de 1,2 milhões de palavras: as revistas portuguesas (14) *Activa*, *Caras Especial Moda*, *Cosmopolitan*, *Crónica Feminina*, *Elle*, *Eva*, *Flama*, *GQ*, *Máxima*, *Maxmen*, *Men's*, *Health*, *Modas e Bordados*, *Técnicas de Alfaiataria*, *Vestir* e *Vogue* e as revistas brasileiras (14) *Boa Forma*, *Burda*, *Cigarra*, *Claudia*, *Cruzeiro*, *Desfile*, *Estilo*, *Figurino*, *Manchete*, *Manequim*, *Marie Claire*, *Máxima*, *Moda e Moldes* e *Vogue*. Contém ainda 1.300 imagens de etiquetas de montras de lojas de vestuário de Lisboa e de Braga e de São Paulo e Rio de Janeiro.

Estes dois sub-corpora integram o corpus CONDIVport, em construção, estruturado na base de três variáveis: (i) geográfica (Portugal vs. Brasil), (ii) diacrónica (1950 – 1970 – 1990/2000) e (iii) estilística (jornais, revistas > etiquetas, Net fóruns > Net *chats*). Actualmente, contém 4 milhões de palavras do registo formal (jornais de desporto e revistas de moda) e 15 milhões de palavras do registo informal (*chats* de futebol e etiquetas de lojas de vestuário). Está parcialmente disponibilizado no sítio da Linguatca, em www.linguatca.pt (projecto AC/DC).

A análise compreende 21 perfis onomasiológicos de futebol, perfazendo um total de 183 termos, e 22 perfis onomasiológicos do vestuário de homem (M) e de mulher (F), perfazendo um total de 264 termos, a seguir identificados (para cada perfil, apresentamos o nome do conceito/referente em versaletes e os sinónimos referenciais seleccionados).⁴

Termos de futebol:

ÁRBITRO: *apitador*, *árbitro*, *director da partida*, *juiz*, *juiz de campo*, "referee", *referi*, *refre*
 ÁRBITRO AUXILIAR: *árbitro auxiliar*, *árbitro assistente*, *auxiliar*, 2º/3º/4º *árbitro*, *bandeirinha*, *fiscal de linha*, *juiz de linha*, "liner"
 AVANÇADO: *atacante*, *avanzado*, *avante*, *dianteiro*, "forward", *ponta-de-lança*
 BALIZA: *arco*, *baliza*, *cidadela*, "goal", *gol*, *malhas*, *marco*, *meta*, *rede*, *redes*, *vala*
 BOLA: *balão*, *bola*, *couro(inho)*, *esfera*, *esférico*, *pelota*
 DEFESA: "(full-)back", *beque*, *bequeira*, *defensor*, *defesa*, *lateral*, *libero*, *zagueiro*
 EQUIPA: *conjunto*, *formação*, *eieven*, *equipa/e*, *escrete*, *esquadra*, *esquadrão*, *grupo*, "match", *onze*, *onzena*, *plantel*, *quadro*, "scratch(men)", "team", *time*, *turma*

⁴ As duas listas de termos foram elaboradas a partir dos dicionários de referência da língua portuguesa (Academia das Ciências de Lisboa, Houaiss, Aurélio, Porto Editora), de dicionários e glossários de futebol, dos estudos linguísticos de Feijó (1994, 1998) sobre a linguagem do futebol, do glossário de moda de Farias (2001, 2003) e naturalmente também com base nos corpora construídos. Os estrangeirismos que conservam a forma original estão indicados entre aspas. Nem todas as variantes gráficas estão aqui indicadas. Foram excluídos os termos ou outras expressões marcadamente populares.

- EXTREMO: *ala, extremo, ponta, ponteiro*
 FALTA: *carga, falta, "foul", golpe (baixo, irregular), infra(c)ção, obstru(c)ção, transgressão, violação (das regras)*
 FINTA: *corte, dribble(ing), engano, "feint", finta, firula, ginga, lesa, manobra enganadora, simulação*
 FORA-DE-JOGO: *adiantamento, banheira, deslocação, fora-de-jogo, impedimento, "offside", posição irregular*
 GOLO: *bola, "goal", gol, golo, ponto, tento*
 GRANDE PENALIDADE: *castigo máximo, castigo-mor, falta máxima, grande penalidade, penalidade, penalidade máxima, penákti/pênalti, "penalty"*
 GUARDA-REDES: *arquero, "goal-keeper", goleiro, golquiper, guarda-meta, guarda-redes, guarda-vala(s), guardião, "keeper", porteiro, quiper, vigia*
 JOGADA: *jogada, lance*
 JOGO: *batalha, choque, combate, competição, confronto, desafio, disputa, duelo, embate, encontro, jogo, justa, luta, "match", partida, peladinha, peleja, prêlio, prova, pugna*
 MÉDIO: *alfe, central, centro-campista, centro-médio, "half", interior, médio, meia, meio-campista, meio-campo, "midfield", trinco, volante*
 PONTAPÉ: *chute, chuto, "kick(-off)", panázio, pelotada, pontapé, quique, "shoot", tiro*
 PONTAPÉ DE CANTO: *canto, chute de canto, "corner", córner, escanteio, esquinado, pontapé de canto, tiro de canto*
 PONTAPÉ LIVRE: *chute (in)direto, falta, "free(-kick)", livre (directo, indirecto), pontapé livre, tiro dire(c)to, tiro livre (direto, indirecto)*
 TREINADOR: *mister, professor, técnico, treinador*

Termos de vestuário:

- BLUSA F: *"blouse", blusa, blusinha, "bustier", camisa, camisa-body, camisão, camiseiro(inho), camiseta/e, (blusa) "chémisier", (blusa) chemisiê*
 BLUSÃO M/F: *"blazer", blêizer, blusão, "bluson", camurça, camurcine, camisa esporte, casaco de pele, ganga, etc., colete, parka*
 CALÇAS M/F: *calça, calças, pantaloná*
 CALÇAS CURTAS M/F: *bermuda(s), calças-capri, calça(s) corsário, calça(s) curta(s), calças 3/4, calções, "cool pants", corsários, "hot pants", "knickers", "pantacourt", "pedal pusher", "short(s)", "short cuts", "short shorts", shortinho, "slack(s)"*
 CALÇAS JUSTAS F: *"fuseau(x)", fusô, "legging(s)"*
 CAMISA M: *blusão, camisa, camisa de gravata, camisa de manga curta, camisa desportiva, camisa esporte(iva), camisa jeans, camisa social, camiseta, camiseta, "camisette", camisinha*
 CAMISOLA M/F: *blusa, blusão, blusinha, "body", cachemir, camisa, camisa-de-meia, camiseta, camisinha, camisola, camisolinha, "canoutier", canoutiê, malha, malhinha, moleton, "pull", "pullover", pulóver, suéter, "sweat", "sweat shirt", "sweater"*
 CASACO F: *"blazer", blêizer, casaco, casaquinho/a, "manteau", mantô, paletó, "paletot"*
 CASACO M: *"blazer", blêizer, casaco, paletó, "paletot"*
 CASACO CURTO F: *bolero, carmona, casa(i)b(v)equê, casaco curto, casaquilha, colete, colete camiseiro, corpete, corpinho, garibáldi, "gilet", manguito, mini, minicasaco, roupinha, "shortie", vasquinha*
 CASACO CURTO M: *casaco curto, colete, espartilho, gibão, "gilet", jaleca, jaleco, jaqueta, véstia*
 CASACO DE CERIMÓNIA M/F: *"black-tie", casaca, casaco cerimónia, fraque, "manteau", mantô, paletó, "paletot", "pelerine", "smo(c)king", sobrecasaca, "tuxedo"*
 CASACO DE MALHA M/F: *cardigã, "cardigan", casaco/casaquinho de malha (de lã, de tricó), "gilet", japona, malha, "twin-set"*
 CASACO IMPERMEÁVEL M/F: *"ciré", "ciré-maxi", "anorak", canadiana, capa, capa de chuva, casaco impermeável, corta-vento, casaco-gabardina, gabardine/a, impermeável, kispo, parka, redingote*
 CASACO QUENTE (Inverno) M/F: *abafo, agasalho, balandrau, capote, casacão, casaco comprido, casaco de abafo/abafar, casaco de agasalho, casaco de/em pele, casaco-sobretudo, "duffle-coat",*

gabão, "gilet", "manteau", mantô, manto, overcoat, paletô, "pardessus", "pelerine", samarra, sobrecasaca, sobretudo, sobreveste, "trench (coat)"
 CONJUNTO F: *"complet", completo, conjunto, costume, duas-peças, "ensemble", fatinho, fato, saia-casaco, "tailleur", "toilette", toilette, vestido-casaco*
 FATO M: *beca, completo, costume, fato, terno*
 JAQUETA M/F: *casaca, casaco curto, jaleca, jaqueta, "jaquette", jaquetinha, véstia*
 JEANS M/F: *calça(s) de ganga, calça(s) em denim, calça(s) em jeans, ganga, jeans*
 SAIA F: *kilt, maxi (máxi), maxissaia, micro-mini, micro-saia, mini (mini), mini-saia, minissaia, pareô, saia, saia-calça, saia-calção, saião, sainha, saíote*
 T-SHIRT M/F: *camisa, camiseta/e, "camisette", camisola, licra, "singlet", "tee-shirt", "t-shirt"*
 VESTIDO F: *camiseiro, "chemisier", chemisiê, "shirt-dress", traje/o, veste, vestido(inho), vestido-camisa, vestido-camiseiro, vestido-camiseta, vestido-chemiser(ê), (vestido) cai-cai, (vestido) tomara-que-caia*

A base de dados de futebol é constituída por 90.202 observações do uso dos referidos termos nos jornais de desporto dos primeiros anos das décadas de 50, 70 e 90-2000 acima indicados, e por 143.946 observações do uso dos mesmos termos nos *chats* de futebol. A base de dados de moda é constituída por 12.448 observações do uso dos referidos termos de vestuário numa parte das revistas de moda acima indicadas⁵ e por 3.240 observações do uso dos mesmos termos nas 1.300 etiquetas de montras de lojas de vestuário acima referidas.

Pelas limitações de espaço, apresentamos a seguir os resultados da pesquisa de modo muito sumário (para uma análise detalhada, incluindo as frequências dos termos de futebol e vestuário observados, ver Silva & Marlene, 2005 e Silva, 2006).

5. Resultados: termos de futebol

Começamos pela questão central da convergência/divergência. O Quadro 1 apresenta os valores de uniformidade de cada conceito (U) e de uniformidade média (U') dos 21 perfis onomasiológicos de futebol do português de Portugal (P) e do Brasil (B), dos primeiros anos das décadas de 50 (P50/B50), 70 (P70/B70) e 90-2000 (P00/B00); na coluna da direita, os totais de ocorrências dos conceitos e termos analisados. Lembremos que a medida U responde à questão "qual é a evolução (convergente ou divergente) se tomarmos cada conceito tão importante quanto o outro?" e a medida U' responde à questão "qual é a evolução (convergente ou divergente) se tivermos em conta a frequência relativa dos diferentes conceitos?". É portanto a medida U' a que melhor indica a situação da amostra. Neste cálculo não separámos as variantes *golo* (P) e *gol* (B) nem *chuto* (P) e *chute* (B). Juntando, em cada perfil, os termos formados do mesmo radical mas com diferenças de pronúncia, os resultados são idênticos.

⁵ Designadamente as revistas *Modas e Bordados, Técnicas de Alfaiataria e Vestir* de P50 e *Cruzeiro e Estado* do B50; *Modas e Bordados* de P70 e *Cláudia, Figurino e Manequim* de B70; *Activa, Caras Especial Moda, Cosmopolitan, Elle, GQ, Máxima, Maxmen, Men's Health* e *Vogue* de P00 e *Boa Forma, Cláudia, Estilo, Manequim* e *Vogue* de B00.

	P50/B50		P70/B70		P00/B00		Total n°
	U	U'	U	U'	U	U'	
ÁRBITRO	43,9	1,76	22,0	0,82	90,2	2,85	3310
ÁRBITRO AUXILIAR	20,1	0,09	43,4	0,33	18,2	0,18	623
AVANÇADO	16,9	0,65	28,8	0,82	10,1	0,39	3238
BALIZA	28,9	1,35	12,2	0,42	7,5	0,17	3250
BOLA	61,2	4,99	81,9	7,04	95,0	4,75	6542
DEFESA	19,9	0,48	15,2	0,5	34,3	1,33	2791
EQUIPA	37,3	7,19	33,5	7,04	41,8	7,83	17642
EXTREMO	6,5	0,14	2,6	0,04	29,3	0,08	1200
FALTA	60,2	0,61	93,0	0,64	92,7	0,85	814
FINTA	73,1	0,22	67,1	0,23	54,5	0,11	251
FORA-DE-JOGO	2,8	0,01	0,0	0	0,0	0	395
GOLO	42,5	4,79	93,8	12,2	94,1	12,9	11294
GRANDE PENALIDADE	27,3	0,41	1,9	0,02	0,6	0,01	1450
GUARDA-REDES	8,0	0,21	1,0	0,03	0,0	0	2332
JOGADA	91,3	3,97	69,4	3,84	52,7	2,2	4140
JOGO	54,6	14,1	75,5	17,4	73,0	15,9	21502
MÉDIO	48,3	1,22	19,7	0,13	0,9	0,03	2004
PONTAPÉ	47,8	0,9	18,6	0,23	6,4	0,05	1211
PONTAPÉ DE CANTO	0,5	0,01	0,0	0	0,0	0	818
PONTAPÉ LIVRE	0,0	0	1,0	0,01	6,1	0,06	675
TREINADOR	49,8	0,69	73,5	3,43	65,4	7,05	4720
Total	35,3	43,8	35,9	55,2	36,8	56,8	90202

Quadro 1: Percentagens de uniformidade U e U' dos 21 conceitos de futebol

Os totais de U' indicam portanto convergência, maior de 50 para 70 e muito pequena de 70 para 2000:

$$\begin{array}{rcl}
 U' (P50,B50) & < & U' (P70,B70) \\
 43,8\% & < & 55,2\% \\
 & & \equiv \\
 & & U' (P00,B00) \\
 & & 56,8\%
 \end{array}$$

U' aumenta 11,4% de 50 para 70 e somente 1,6% de 70 para 2000. Quer isto dizer que a hipótese da divergência não se confirma. Todavia, os resultados mostram também que a distância lexical entre as duas variedades é relativamente grande: a uniformidade situa-se entre 44% e 57%, isto é, metade da amostra são diferenças, divergências.

Tentemos interpretar os resultados obtidos, no sentido de verificar possíveis correlações entre eles e determinadas propriedades dos termos em análise. Por exemplo, quais são os perfis que se comportam diferentemente da tendência convergente global? Terão os respectivos conceitos algo em comum? Os perfis com resultados divergentes

são 'baliza', 'finta', 'jogada', 'médio', 'pontapé' e, em grau menor, 'avançado', 'fora-de-jogo', 'grande penalidade' e 'guarda-redes'. Mas não apresentam características que os tornem num subgrupo distinto.

Mais importante é saber qual a influência dos estrangeirismos, principalmente anglicismos, na tendência convergente global. O peso de termos do Inglês diminui ou aumenta? É maior no PE ou no PB?

Procedemos agora a um novo cálculo, que consiste em medir a proporção de termos com a característica (A) 'anglicismo' ou 'estrangeirismo' (qualquer que seja a proveniência) no perfil onomasiológico de um conceito e, depois, no conjunto dos 21 conceitos analisados, nas amostras do PE e PB. É importante notar que neste cálculo a atribuição da característica A não é uma questão binária (presença ou ausência), mas um *continuum*: é atribuído o valor mais alto (1) aos estrangeirismos que mantêm a forma original e o valor mais baixo (0,25, por exemplo) a fortes adaptações e a decalques.

O Quadro 2 apresenta os resultados com as percentagens distribuídas por dois grupos: anglicismos (A'_{ing}) e todos os estrangeirismos (A'_{estrang}).

A'_{ing} (P50)	7,1%	18,0%	A'_{ing} (B50)
A'_{ing} (P70)	9,8%	17,1%	A'_{ing} (B70)
A'_{ing} (P00)	10,2%	16,2%	A'_{ing} (B00)
A'_{estrang} (P50)	13,9%	23,5%	A'_{estrang} (B50)
A'_{estrang} (P70)	17,9%	22,8%	A'_{estrang} (B70)
A'_{estrang} (P00)	18,5%	23,3%	A'_{estrang} (B00)

Quadro 2: Percentagens de estrangeirismos

A influência de anglicismos e outros estrangeirismos é claramente maior no PB do que no PE. A grande diferença percentual entre B50 e P50 resulta do facto de haver um maior número e frequência de estrangeirismos conservando a sua forma original no PB. Estes resultados contribuem, em parte, para a distância maior entre as duas variedades nos anos 50 e, conseqüentemente, para a tendência convergente global.

Um outro dado interessante tem a ver com a tendência à adaptação dos termos estrangeiros ou à sua substituição por termos vernáculos. Medindo agora a percentagem de adaptações, desde a adaptação mais próxima do termo estrangeiro à mais afastada, o resultado é:

$A'_{\text{ing.adapt}}$ (P50)	6,0%	2,8%	$A'_{\text{ing.adapt}}$ (B50)
$A'_{\text{ing.adapt}}$ (P70)	7,9%	16,9%	$A'_{\text{ing.adapt}}$ (B70)
$A'_{\text{ing.adapt}}$ (P00)	8,9%	16,0%	$A'_{\text{ing.adapt}}$ (B00)

Quadro 3: Percentagens de adaptações/decalques de anglicismos

A percentagem de adaptações de termos ingleses aumenta fortemente no PB de 50 para 70. No conjunto dos 21 perfis, temos 23 adaptações e 19 decalques no PB contra 6 adaptações e 14 decalques do PE.

Tudo isto vem confirmar a maior tendência do PB, não só para a importação directa de estrangeirismos, como também para a sua adaptação ou decalque; pelo contrário, no PE os estrangeirismos tendem a ser substituídos por termos vernáculos. Alguns exemplos: *goleiro* (PB) e *guarda-redes* (PE), *gol* (PB) e *baliza* (PE), *córner/escanteio* (PB) e *pontapé de canto* (PE), *pênalty* (PB) e *grande penalidade* (PE), *beque* (PB) e *defesa* (PE) e, como excepção, *impedimento* (PB) e *fora-de-jogo* (PE).

Uma outra questão é a da esperada influência do PB no PE, justamente neste domínio do futebol. O Quadro 4 sumaria os resultados percentuais de toda a amostra, em dois cálculos: as percentagens da esquerda têm em conta apenas os brasileirismos de que há menos dúvidas e as da direita, todos os brasileirismos que estão atestados nalgum dicionário de referência.

A'_{Braz} (P50)	0,8%	2,3%
A'_{Braz} (P70)	1,0%	3,4%
A'_{Braz} (P00)	1,1%	2,0%

Quadro 4: Percentagens de brasileirismos

Os resultados não são concludentes: não permitem confirmar claramente a hipótese da influência do PB no PE.

Passemos à questão da estratificação ou distância actual entre registo padrão e registo subpadrão (de nível intermédio). Para o efeito, calculámos a uniformidade, em cada uma das variedades, entre os dados dos jornais de desporto dos anos 90/2000 (P00 e B00) e os dados dos *chats* de clubes de futebol ($P_{\text{sub}00}$ e $B_{\text{sub}00}$). Os resultados são os seguintes:

$$\begin{array}{l} U' (P00, P_{\text{sub}00}) \\ 80,9\% \end{array} \quad \begin{array}{l} \cong \\ \cong \end{array} \quad \begin{array}{l} U' (B00, B_{\text{sub}00}) \\ 78,8\% \end{array}$$

Os resultados mostram um grau relativamente elevado de uniformidade entre padrão e subpadrão dentro de cada uma das variedades (a uniformidade entre as duas variedades a nível do estrato subpadrão desce para 59,2%, situando-se num valor muito próximo do da uniformidade a nível do estrato padrão, que é, como vimos, de 56,8%). Mais significativamente, estes resultados não permitem confirmar a hipótese de uma distância maior entre padrão e subpadrão no PB.

6. Resultados: termos de vestuário

O Quadro 5 apresenta os valores de uniformidade U e U' dos 22 perfis onomasiológicos de vestuário estudados.

CONVERGÊNCIA E DIVERGÊNCIA NO LÉXICO DO PORTUGUÊS EUROPEU

	P50/B50		P70/B70		P00/B00		Total n°
	U	U'	U	U'	U	U'	
BLUSA F	90,6	9,17	83,1	8,39	85,3	7,82	1213
BLUSÃO M/F	15,4	0,07	98,4	1,78	78,1	1,6	193
CALÇAS M/F	100	3,09	80,3	9,25	97,4	14,6	1327
CALÇAS CURTAS M/F	68,2	1,31	67,9	2,58	19,9	0,93	454
CALÇAS JUSTAS F	0	0	0	0	81,8	0,2	12
CAMISA M	100	1,33	83,3	0,74	97,6	0,93	128
CAMISOLA M/F	50,5	0,98	38,8	2,39	12,7	0,63	573
CASACO F	89,6	4,59	66,9	4,82	48	4,12	896
CASACO M	46,8	1,18	5,13	0,05	22,2	0,06	138
CASACO CURTO F	69,3	1,3	66	1,18	79,2	0,92	195
CASACO CURTO M	85	0,55	0	0	100	0,12	39
CASACO DE CERIMÓNIA M/F	86	2,07	80	0,13	56,3	0,24	106
CASACO DE MALHA M/F	100	0,49	14,3	0,12	17	0,15	97
CASACO IMPERMEÁVEL M/F	60,3	1,25	68,1	1,6	34,3	0,26	207
CASACO QUENTE M/F	40,3	2,71	57,1	3,37	71,8	2,47	642
CONJUNTO F	63,2	6,76	79,3	5,94	23,9	0,86	849
FATO M	0	0	0	0	0	0	88
JAQUETA M/F	48	0,68	88,7	1,91	35,3	0,42	198
JEANS M/F	0	0	0	0	79,4	2,89	177
SAIA F	99,8	21,9	90,5	16,2	91,7	16	2337
T-SHIRT M/F	0	0	77,1	1,16	17,7	0,59	244
VESTIDO F	98,7	23,6	92,7	15,2	92,4	16,1	2335
Total	59,6	82,9	56,3	76,9	56,5	71,7	12448

Quadro 5: Percentagens de uniformidade U e U' dos 22 conceitos de vestuário

Os dados indicam portanto divergência:

$$\begin{array}{rcl}
 U' (P50,B50) & > & U' (P70,B70) & > & U' (P00,B00) \\
 82,9\% & > & 76,9\% & > & 71,7\%
 \end{array}$$

U' diminui 6% de 50 para 70 e 5,2% de 70 para 2000, isto é, U' diminui 11,2% de 50 para 2000 (22,1% se separarmos todas as variantes fonéticas e gráficas, os diminutivos e as expressões abreviadas). A hipótese da divergência entre PE e PB é pois confirmada com estes resultados do vocabulário do vestuário.

Tentemos interpretar os resultados obtidos, no sentido de verificar possíveis correlações entre eles e determinadas propriedades dos termos em análise. Os perfis com resultados claramente contrários (convergentes) são 'blusão' e 'casaco quente'.

Mas não apresentam características que os tornem num subgrupo distinto. Mais importante é saber qual a influência dos estrangeirismos na tendência divergente global.

O Quadro 6 apresenta os resultados com as percentagens distribuídas por três grupos: francesismos (A'_{Fr}), anglicismos (A'_{Ing}) e todos os estrangeirismos (A'_{estrag}).

A'_{Fr} (P50)	17,6%	18,5%	A'_{Fr} (B50)
A'_{Fr} (P70)	15,9%	18,1%	A'_{Fr} (B70)
A'_{Fr} (P00)	10,2%	7,9%	A'_{Fr} (B00)
A'_{Ing} (P50)	3,3%	4,2%	A'_{Ing} (B50)
A'_{Ing} (P70)	5,8%	7,6%	A'_{Ing} (B70)
A'_{Ing} (P00)	16,9%	17,0%	A'_{Ing} (B00)
A'_{estrag} (P50)	22,4%	23,8%	A'_{estrag} (B50)
A'_{estrag} (P70)	22,1%	26,7%	A'_{estrag} (B70)
A'_{estrag} (P00)	28,2%	24,9%	A'_{estrag} (B00)

Quadro 6: Percentagens de estrangeirismos

Há tanto no PE como no PB uma significativa diminuição de francesismos (mais acentuada no PB) e um claro aumento de anglicismos, por razões sócio-culturais bem conhecidas. No conjunto de todos os estrangeirismos, temos um aumento no PE e uma situação de estabilidade no PB. A influência dos estrangeirismos é maior no PB nos anos 50 e 70, ao passo que nos últimos anos é um pouco maior no PE. Este resultado confirma a hipótese de uma maior permeabilidade do PB aos estrangeirismos, quer importando-os directamente, quer adaptando-os. Todavia, não sabemos ainda qual o seu efeito na tendência divergente global.

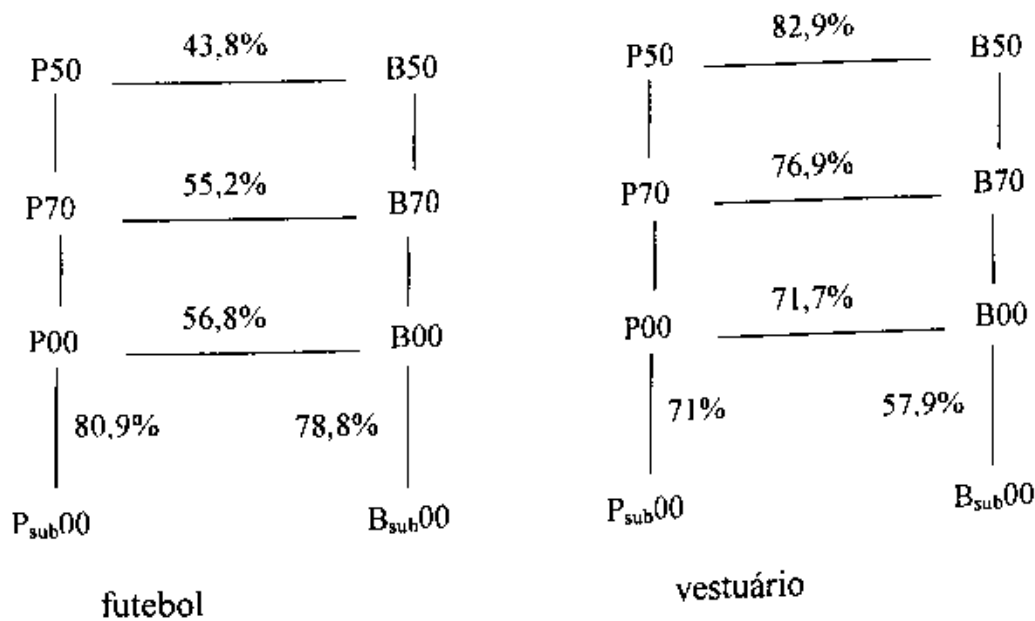
Passemos à questão da estratificação ou distância actual entre registo padrão e registo subpadrão. Para o efeito, calculámos a uniformidade, em cada uma das variedades, entre os dados do registo mais formal e mais nacional das revistas de moda dos anos 90/2000 (P00 e B00) e os dados do registo menos formal e mais local das lojas de vestuário (P_{sub00} e B_{sub00}). Os resultados são os seguintes:

$$\begin{array}{lcl} U' (P00, P_{sub00}) & > & U' (B00, B_{sub00}) \\ 71\% & > & 57,9\% \end{array}$$

A distância entre padrão e subpadrão é pois maior no PB do que no PE, confirmando assim a hipótese sobre esta questão. Por outro lado, a distância entre as duas variedades é claramente maior a nível do estrato subpadrão (45,8%) do que a nível do estrato padrão (71,7%, indicado acima).

7. Conclusão e trabalho futuro

O Quadro 7 sintetiza os resultados relativamente às duas questões da presente investigação: convergência/divergência e estratificação.



Quadro 7: Valores globais de uniformidade diacrónica entre PE e PB e de uniformidade actual entre padrão e subpadrão

Os resultados dos termos de futebol e dos termos de vestuário são diferentes, sobretudo em relação à questão principal da convergência ou divergência entre PE e PB. Os termos de vestuário confirmam a hipótese de divergência e a hipótese de uma distância maior entre padrão e subpadrão (estrato intermédio) na variedade brasileira. Pelo contrário, os termos de futebol apontam no sentido da convergência, mais clara de 50 para 70 e pouco significativa de 70 para 2000, mas ao mesmo tempo mostram uma uniformidade consideravelmente menor (distância maior) entre as duas variedades em todos os períodos. Esta disparidade de resultados é um motivo acrescido para a análise de outros campos lexicais.

No plano teórico e metodológico, esperamos ter conseguido mostrar as vantagens da perspectiva cognitiva e de métodos quantitativos no estudo das relações entre variedades nacionais (e outras variedades *lectais*) de uma língua e, assim, a inevitabilidade da sociolinguística (e, por extensão, da sociolinguística) cognitiva.

Naturalmente que mais trabalho há a fazer. Como extensões actuais, vamos incluir (i) mais termos de futebol e de vestuário, particularmente de outros subdomínios da moda (tecido, padrão e estilo); (ii) dados atitudinais (de um inquérito a falantes dos dois países); e, mais importante ainda, (iii) outros campos lexicais (está em construção um *corpus* de saúde), bem como conceitos isolados. Futuramente, esperamos poder incluir (i) palavras funcionais (preposições, por exemplo) e variáveis não-lexicais (morfológicas e sintácticas) e (ii) técnicas de análise multivariacional.

Referências

- Belard, Francisco (2001) Brasil e Portugal. *Jornal Expresso*, 5 Maio 2001, Cartaz, p. 66.
- Caetano, José A. Palma (2002) Algumas tendências actuais da evolução da língua portuguesa em Portugal. In: Bernhard Pöll & Franz Rainer (eds.), *Vocabula et Vocabularia: Études de Lexicologie et de (Méta-)Lexicographie Romanes en l'Honneur du 60^e Anniversaire de Dieter Messner*. Frankfurt: Lang, pp. 53-65.
- Castro, Ivo (1986) Um juízo sobre o novo Acordo Ortográfico. *ICALP* 5, pp. 41-48.
- Farias, Emília Faria Peixoto (2001) *A Linguagem da Moda no Português Contemporâneo*. Dissertação de doutoramento. Recife: UFPE.
- Farias, Emília Faria Peixoto (2003) *Glossário de termos da Moda*. Fortaleza: Editora UFC / Edição Sebrae/CE.
- Feijó, Luiz Cesar Saraiva (1994) *A linguagem dos esportes de massa e a gíria no futebol*. Rio de Janeiro: UERJ-Tempo Brasileiro.
- Feijó, Luiz Cesar Saraiva (1998) *Brasil x Portugal: Um derby linguístico*. Rio de Janeiro: Sociedade Brasileira de Língua e Literatura.
- Geeraerts, Dirk (2001) On measuring lexical convergence. In: Augusto Soares da Silva (org.), *Linguagem e Cognição: A Perspectiva da Linguística Cognitiva*. Braga: Associação Portuguesa de Linguística e Universidade Católica Portuguesa, pp. 51-61.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman (1999) *Convergentie en divergentie in de Nederlandse woordenschat*. Amsterdam: Meertens Instituut.
- Saraiva, José António (1986) O novo Português. Editorial do *Jornal Expresso*, 24 de Maio, p. 2.
- Silva, Augusto Soares da (2004a) Léxico e variação Portugal/Brasil: Para uma sociolinguística cognitiva do Português. *Revista Portuguesa de Humanidades* 8, pp. 99-117.
- Silva, Augusto Soares da (2004b) Palavras e conceitos no tempo: para uma onomasiologia diacrónica e cognitiva do Português. In: Graça Rio-Torto et al. (orgs.), *Dar a Palavra à Língua. Homenagem a Mário Vilela*. Porto: Faculdade de Letras do Porto, no prelo.
- Silva, Augusto Soares da (2004c) Introdução: linguagem, cultura e cognição, ou a Linguística Cognitiva. In: Augusto Soares da Silva, Amadeu Torres & Miguel Gonçalves (orgs.), *Linguagem, Cultura e Cognição: Estudos de Linguística Cognitiva*. Vol. 1. Coimbra: Almedina, pp. 1-18.
- Silva, Augusto Soares da (2005) Para o estudo das relações lexicais entre o Português Europeu e o Português do Brasil: Elementos de sociolinguística cognitiva e quantitativa do Português. In: Inês Duarte & Isabel Leiria (orgs.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*, pp. 211-226.
- Silva, Augusto Soares da (2006) O léxico do futebol no Português Europeu e no Português Brasileiro: convergência ou divergência. *Diacrítica* 20 (1), no prelo.
- Silva, Augusto Soares da & Marlene Danaia Duarte (2005) O léxico do vestuário no Português Europeu e no Português Brasileiro: convergência ou divergência? *Revista Portuguesa de Humanidades* 9, pp. 117-136.
- Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts (2003) Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37, pp. 317-337.