

# MEDINDO A CONVERGÊNCIA E DIVERGÊNCIA LEXICAL ENTRE O PORTUGUÊS EUROPEU E O PORTUGUÊS BRASILEIRO

*Augusto Soares da Silva*

*Universidade Católica Portuguesa – Braga*

## INTRODUÇÃO

Este estudo apresenta os instrumentos e os resultados principais de duas questões de investigação sobre as relações lexicais entre o Português Europeu (PE) e o Português Brasileiro (PB): (i) saber se as duas variedades se encontram num processo de convergência ou de divergência a nível do léxico nos últimos 50 anos; e (ii) saber que influência têm factores linguísticos internos na tendência evolutiva global.<sup>1</sup> Ambas as questões implicam uma análise diacrónica, sendo *externa* a primeira e *interna* a segunda. O objecto de análise é a variação onomasiológica que envolve sinónimos *de-*

---

<sup>1</sup> Na base deste estudo está o projecto de investigação “Convergência e Divergência no Léxico do Português”, financiado pela Fundação para a Ciência e a Tecnologia de Portugal (Ref<sup>o</sup> POCTI/LIN/48575/2002), já concluído (2004-2006). Outras questões do projecto CONDIV são sincrónicas: estratificação actual das duas variedades e atitudes dos falantes dos dois países. Para mais informações, ver <[www.facfil.ucp.pt/condiv.html](http://www.facfil.ucp.pt/condiv.html)>.

*notacionais* e a base empírica consiste em largos milhares de observações do uso de palavras alternativas designando conceitos dos domínios do futebol e do vestuário. Esta investigação sociolinguística situa-se num quadro cognitivo (da Linguística Cognitiva, de R. Langacker, G. Lakoff & L. Talmy) e quantitativo (pela utilização de métodos quantitativos de medição da convergência e divergência). Num primeiro momento, exporemos os elementos essenciais do enquadramento teórico e metodológico deste estudo de sociolinguística cognitiva e quantitativa. A seguir, apresentaremos os resultados da análise diacrónica externa e interna, bem como da correlação dos factores externos e internos.

Como hipóteses sobre as relações entre PE e PB, admite-se (i) uma influência crescente do PB sobre o PE, (ii) maior receptividade do PB aos estrangeirismos e (iii), embora não se encontrem hipóteses claras na literatura, a ideia generalizada de uma fragmentação progressiva e inelutável e, portanto, a hipótese da divergência.

## 1. ELEMENTOS TEÓRICOS E METODOLÓGICOS

Três razões justificam a opção pelo quadro teórico da Linguística Cognitiva: (i) a sua orientação *recontextualizante* (reintegrando as diferentes formas de *contexto*, excluídas pelos modelos gramaticais autonomistas, particularmente o generativista), para o *significado* (incluindo o significado social) e *baseada-no-uso* (origem da própria variação linguística); (ii) a importância dada à flexibilidade e à variação, temas que estão na base da teoria do protótipo e da teoria da metáfora conceptual; e (iii) a assunção de que não só a capacidade para a linguagem se fundamenta em capacidades cognitivas gerais, como também todas estas capacidades são cultural e socialmente situadas e definidas (donde a noção de *cognição social*). Mais especificamente, a nossa investigação insere-se no âmbito da emergente Sociolinguística Cognitiva (Geeraerts 2005, Kristiansen & Dirven 2007), inevitavelmente implicada por aqueles princípios, que tem desenvolvido métodos quantitativos e multivariacionais para a análise da confluência de factores conceptuais, discursivos e variacionais (ver também Silva 2006b, 2007). Neste âmbito, a presente investigação apoia-se na concepção geral e nos métodos quantitativos da investigação sociolinguística desenvolvida por Dirk Geeraerts e sua equipa para o Neerlandês da Holanda e da Bélgica (Geeraerts, Grondelaers & Speelman 1999).

O objecto de investigação é a variação onomasiológica *formal*: aquela que envolve sinónimos denotacionais (ou referenciais), como *avançado*, *atacante* e *dianteiro*, e não a variação onomasiológica *conceptual*, a qual envolve diferenças conceptuais, como *avançado* e *jogador*. Os sinónimos denotacionais evidenciam tipos de diferenças sociolinguísticas e são estas diferenças regionais, sociais, estilísticas, pragmático-discursivas e históricas que definem a própria existência e a competição de variedades *lectais* de uma língua.

Para medir a convergência e divergência e outras distâncias entre o PE e o PB, utilizamos três métodos quantitativos: (i) medida de uniformidade (U); (ii) medida da proporção de determinado traço (A), ligado a itens ou a conceitos lexicais; e (iii) medida da correlação de diferentes variáveis.

Mais importante, a medida U (“uniformidade linguística baseada em perfis”) envolve duas noções: *perfil onomasiológico* ou conjunto de sinónimos usados para designar um conceito/referente, diferenciados pela sua frequência relativa, e *uniformidade* ou 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 de A e 7 de B, resulta do número de pares comuns de nomeação desse conceito (7 pares), sendo portanto  $U = 70\%$ . Este mesmo resultado obtém-se somando as frequências relativas mais pequenas de cada termo alternativo: 30% de A e 40% de B.

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

$$U_Z(Y_1, Y_2) = \sum_{i=1}^n \min(F_{Z,Y_1}(x_i), F_{Z,Y_2}(x_i))$$

Isto é, a uniformidade U para um conceito Z entre duas amostras  $Y_1$  e  $Y_2$  equivale à soma 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:

$$U^*(Y_1, Y_2) = \sum_{i=1}^n U_{Z_i}(Y_1, Y_2) \cdot G_{Z_i}$$

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  $Z_n$ . Porque toma em conta a frequência relativa, a medida  $U'$  é mais representativa do que a medida  $U$ .

A proporção de termos com determinado traço (medida  $A$ ) é calculada pelas seguintes duas fórmulas: tal como para os cálculos de  $U$  e  $U'$ , a primeira ( $A$ ) toma cada conceito tão importante quanto o outro, ao passo que a segunda ( $A'$ ) tem em conta a frequência relativa de cada conceito.

$$A_{K,Z}(Y) = \sum_{i=1}^n F_{Z,Y}(X_i) \cdot W_{X_i}(K)$$

A proporção  $A$  de todos os termos  $x$  com o traço  $K$  no perfil onomasiológico do conceito  $Z$  na amostra  $Y$  é igual à soma das frequências relativas dos termos  $x$  ponderada pelo valor de pertença a  $W$ .

$$A'_K(Y) = \sum_{i=1}^n A_{K,Z_i}(Y) \cdot G_{Z_i}(Y)$$

A proporção  $A'$  de todos os itens  $x$  com o traço  $K$  na amostra  $Y$  é igual à soma de todas as medidas- $A$  ponderada por  $G$ , isto é, a frequência relativa do conceito  $Z$  em  $Y$ .

A terceira medida, da correlação de diferentes variáveis, é calculada pelo método estatístico da análise de regressão (linear e multilinear).

Os dados pertencem aos campos lexicais 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. Está ainda em análise um terceiro campo lexical – o dos termos de saúde. A base empírica da investigação consiste em largos milhares de observações do uso de termos alternativos para nomear 43 conceitos nominais de futebol e vestuário. Os materiais foram extraídos manualmente de 8 jornais de desporto (*A Bola*, *Record*, *Mundo Desportivo* e *O Jogo*, do

PE; *Jornal dos Sports*, *Gazeta Esportiva*, *Estado de São Paulo* e *Lance*, do PB) e 24 revistas de moda (14 PE e 14 PB) dos primeiros anos das décadas de 50, 70 e 90-2000; e constituem dois sub-*corpora*, com cerca de 4 milhões de palavras, do *corpus* CONDIVport, disponibilizado no sítio da Linguateca, em <[www.linguateca.pt/ACDC](http://www.linguateca.pt/ACDC)>. A análise compreende 21 perfis onomasiológicos de futebol, perfazendo um total de 183 termos, estudados numa base de dados constituída por 90.202 observações do uso destes termos, e 22 perfis onomasiológicos do vestuário de homem (M) e de mulher (F), num total de 264 termos, estudados numa base de dados de 12.451 observações. Para cada perfil, apresentamos o nome do conceito/referente na designação utilizada no PE e os sinónimos denotacionais seleccionados (para não inflacionar as diferenças, foram excluídos os termos marcadamente populares).

Termos de futebol – 21 perfis:

ÁRBITRO: *apitador*, *árbitro*, *director da partida*, *juiz*, *juiz de campo*, “*ref(eree)*”, *referi*, *refre*

ÁRBITRO AUXILIAR: *árbitro auxiliar*, *árbitro assistente*, *auxiliar*, 2º/3º/4º *árbitro*, *bandeirinha*, *fiscal de linha*, *juiz de linha*, “*liner*”

AVANÇADO: *atacante*, *avançado*, *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*, *eleven*, *equipa/e*, *escrete*, *esquadra*, *esquadrão*, *grupo*, “*match*”, *onze*, *onzena*, *plantel*, *quadro*, “*scratch(men)*”, “*team*”, *time*, *turma*

EXTREMO: *ala*, *extremo*, *ponta*, *ponteiro*

FALTA: *carga*, *falta*, “*foul*”, *golpe (baixo, irregular)*, *infra(c)ção*, *obstru(c)ção*, *transgressão*, *violação (das regras)*

FINTA: *corte*, *drible(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, golquíper, guarda-meta, guarda-redes, guarda-vala(s), guardião, “keeper”, porteiro, quíper, vigia*

JOGADA: *jogada, lance*

JOGO: *batalha, choque, combate, competição, confronto, desafio, disputa, duelo, embate, encontro, jogo, justa, luta, “match”, partida, peladinha, peleja, prélio, 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 – 22 perfis:

BLUSA F: *“blouse”, blusa, blusinha, “bustier”, camisa, camisa-body, camisã, 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, pantalone*

CALÇAS CURTAS M/F: *bermuda(s)*, *calças-capri*, *calça(s) corsário*, *calça(s) curta(s)*, *calças 3/4*, *calções*, “*cool pants*”, *corsários*, “*hot pants*”, “*knikers*”, “*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*, *camisete*, “*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*”, *pulôver*, *suéter*, “*sweat*”, “*sweat shirt*”, “*sweater*”

CASACO F: “*blazer*”, *blêizer*, *casaco*, *casquinho/a*, “*manteau*”, *mantô*, *paletó*, “*paletot*”

CASACO M: “*blazer*”, *blêizer*, *casaco*, *paletó*, “*paletot*”

CASACO CURTO F: *bolero*, *carmona*, *casa(i)b(v)eque*, *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/casquinho de malha (de lã, de tricô)*, “*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*, *paletó*, “*pardessus*”, “*pelerine*”, *samarra*, *sobrecasaca*, *sobretudo*, *sobreveste*, “*trench (coat)*”

CONJUNTO F: “*complet*”, *completo*, *conjunto*, *costume*, *duas-peças*, “*ensemble*”, *fatinho*, *fato*, *saia-casaco*, “*tailleur*”, “*toilette*”, *toilete*, *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*, *míni (mini)*, *mini-saia*, *minissaia*, *pareô*, *saia*, *saia-calça*, *saia-calção*, *saião*, *sai-nha*, *saiote*

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*

## 2. ANÁLISE EXTERNA: RESULTADOS DE UNIFORMIDADE

São três as questões específicas:

(1) convergência ou divergência entre o PE e o PB?

(2) a tendência convergente/divergente ocorre dos dois lados ou apenas de um?

(3) a uniformidade interna aumenta ou diminui? é maior no PE ou no PB?

Para as questões (1) e (2), utilizamos os cálculos de uniformidade  $U$  e  $U'$ : convergência e divergência exprimem-se em aumento e diminuição de  $U/U'$ , respectivamente; a evolução convergente ou divergente de uma das variedades em relação à outra manifesta-se em maiores mudanças de  $U/U'$  entre diferentes períodos e em valores de  $U/U'$  maiores ou menores em diferentes períodos do que no mesmo período. Para a questão (3), utilizamos os cálculos de uniformidade interna  $I$  e  $I'$ : um aumento de  $I/I'$  indica homogeneidade interna crescente; a homogeneidade aumenta à medida que



diminuir ou a variação formal (número de termos alternativos) ou a competição entre os termos alternativos (número de termos dominantes). Importa notar que uma diferença inferior a 5% é estatisticamente irrelevante. E, como já foi referido, os cálculos ponderados ( $U'$ ,  $I'$ , etc.) são mais significativos.

O Quadro 1 apresenta os valores de uniformidade de cada conceito/perfil ( $U$ ) e de uniformidade ponderada ( $U'$ ) dos 21 perfis onomasiológicos de futebol do português de Portugal (P) e do Brasil (B) das décadas de 50, 70 e 90-2000; na coluna da direita, o total de observações de cada perfil e o total absoluto de observações.<sup>2</sup>

	P50/B50		P70/B70		P00/B00		Total nº
	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	<b>43,8</b>	35,9	<b>55,2</b>	36,8	<b>56,8</b>	90202

**Quadro 1. Percentagens de  $U$  e  $U'$  dos 21 perfis de futebol**

<sup>2</sup> Por restrições de espaço, não podemos apresentar os resultados para cada um dos referidos termos de futebol (183) e de vestuário (264). Para estes resultados, ver Silva (2006a) e Silva & Duarte (2005).

A Figura 1 sistematiza os resultados percentuais de uniformidade externa e interna dos termos de futebol: em cada par de números, o primeiro é de uniformidade não-ponderada (U, I) e o segundo de uniformidade ponderada (U', I'); os números nas linhas horizontais indicam a percentagem de uniformidade (U/U') entre PE e PB nos anos 50, 70 e 2000, respondendo assim à questão (1); os números nas linhas verticais e diagonais indicam a percentagem de U/U' de período para período e entre diferentes períodos e respondem à questão (2); os números ao lado de cada variedade e período indicam a respectiva percentagem de uniformidade interna (I/I'), respondendo à questão (3).

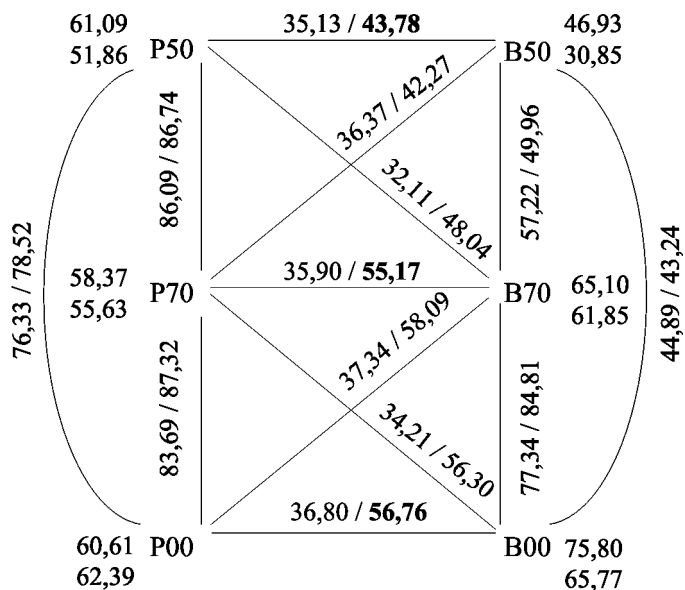


Figura 1. Resultados de uniformidade dos termos de futebol

Em primeiro lugar (questão 1), as percentagens de U' indicam convergência de 50 para 70 (U' aumenta 11,4%) e estabilidade a partir de então, mas as de U não indicam qualquer alteração significativa. Quer isto dizer que é a nível dos conceitos mais frequentes (ver Quadro 1) que aquela convergência se verifica. Ao mesmo tempo, os números mostram uma grande distância entre as duas variedades nos três períodos.

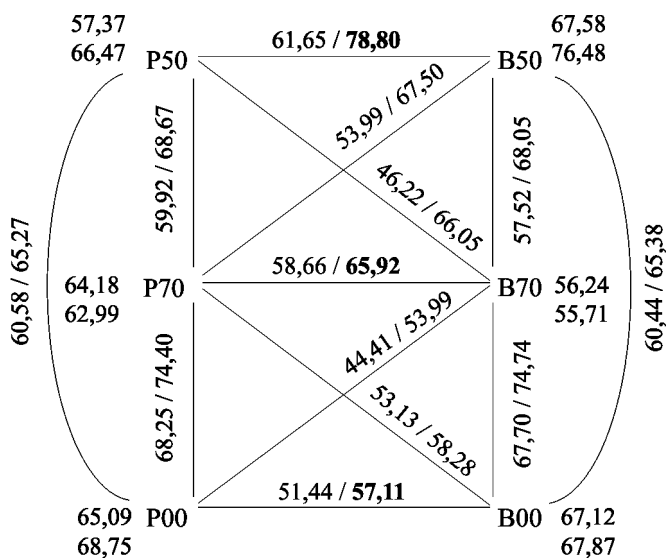
Em segundo lugar (questão 2), as percentagens de U e U' das linhas verticais mostram que há mais mudanças do lado do PB, principalmente a longo prazo (de 50 para 2000) e no primeiro momento (de 50 para 70). Além disso, de 50 para 70 o PB parece aproximar-se do PE (ver a respectiva linha diagonal): B70 está mais próximo, não somente de P70, mas também de P50 (48,04% contra 42,27%). Estes dois resultados coadunam-se com a convergência anteriormente observada. Mais adiante, veremos se se trata efectivamente de uma aproximação do PB ao PE e quais os factores internos que explicam estes resultados percentuais.

Finalmente (questão 3), as percentagens de I e I' indicam um grande aumento de uniformidade interna no PB, sobretudo de 50 para 70, ao passo que as alterações no PE são menores. Isto poderá indicar maior estandarização do vocabulário do futebol no PB nos últimos anos e/ou menor preocupação com o apuramento estilístico da língua nos jornais brasileiros de hoje.

O Quadro 2 apresenta os valores de U/U' dos 22 perfis onomasiológicos do vestuário. E a Figura 2 sistematiza os resultados de U/U' e I/I' relacionados com as três questões referidas.

	P50/B50		P70/B70		P00/B00		Total nº
	U	U'	U	U'	U	U'	
CASACO M	3,9	0,1	5,1	0,1	22,2	0,1	138
CASACO F	78,6	4,0	41,0	3,0	37,5	3,2	896
JAQUETA M/F	36,9	0,5	88,7	1,9	35,3	0,4	198
BLUSÃO M/F	15,4	0,1	98,4	1,8	76,9	1,6	193
CASACO DE MALHA M/F	87,3	0,4	7,1	0,1	15,5	0,1	97
CASACO QUENTE M/F	39,7	2,7	56,7	3,3	71,8	2,5	642
CASACO IMPERMEÁVEL M/F	60,3	1,2	65,1	1,5	34,3	0,3	207
CASACO CURTO F	69,3	1,3	66,0	1,2	79,2	0,9	195
CASACO CURTO M	85,0	0,6	0,0	0,0	100,0	0,1	39
CASACO CERIMÓNIA M/F	79,1	1,9	80,0	0,1	56,3	0,2	106
FATO M	0,0	0,0	0,0	0,0	0,0	0,0	88
CONJUNTO F	61,5	6,6	79,3	5,9	23,9	0,9	849
CAMISOLA M/F	36,8	0,7	34,4	2,1	7,1	0,4	573
CAMISA M	100,0	1,3	83,3	0,7	97,6	0,9	128
BLUSA F	86,9	8,8	79,4	8,0	85,3	7,8	1213
T-SHIRT M/F	0,0	0,0	65,3	1,0	17,7	0,6	244
CALÇAS M/F	66,4	2,0	25,5	2,9	13,6	2,0	1327
JEANS M/F	0,0	0,0	0,0	0,0	77,5	2,9	180
CALÇAS CURTAS M/F	68,2	1,3	67,9	2,6	19,9	0,9	454
CALÇAS JUSTAS F	0,0	0,0	0,0	0,0	81,8	0,2	12
VESTIDO F	96,2	23,0	80,6	13,2	92,4	16,1	2335
SAIA F	99,8	21,9	90,5	16,2	86,1	15,0	2337
Total	61,6	<b>78,4</b>	58,7	<b>65,7</b>	51,4	<b>57,1</b>	12451

**Quadro 2. Percentagens de U e U' dos 22 perfis de vestuário**



**Figura 2. Resultados de uniformidade dos termos de vestuário**

As percentagens de U e U' evidenciam divergência ao longo do tempo: U' diminui 21,69% de 50 para 2000. As percentagens de U e U' das linhas verticais e diagonais indicam simetria nas tendências evolutivas das duas variedades e indicam que a divergência se dá dos dois lados nos dois períodos. As percentagens de I e I' mostram um padrão evolutivo idêntico – diminuição de 50 para 70 e aumento de 70 para 2000 – e maiores mudanças no PB.

Resumindo: (i) temos divergência no vocabulário do vestuário e convergência restringida no vocabulário do futebol: a hipótese da divergência confirma-se no vestuário mas não no futebol; (ii) há mais mudanças internas no PB do que no PE em ambos os campos lexicais; (iii) as mudanças mais fortes dão-se de 50 para 70. Por agora, temos apenas indicações indirectas sobre a direcção das mudanças. Importa verificar possíveis correlações entre os resultados obtidos e determinados traços dos termos em análise.

### 3. ANÁLISE INTERNA: RESULTADOS LIGADOS A TRAÇOS DOS ITENS LEXICAIS

Vamos analisar o impacto de três traços dos termos seleccionados:

(1) endo/exogeneidade: saber se existe uma tendência endógena ou exógena que faça com que uma das variedades se aproxime ou se afaste da outra;

(2) influência estrangeira (estrangeirismos): saber qual o impacto do Inglês e do Francês na tendência convergente/divergente global;

(3) influência normativa: saber qual a consequência da propaganda linguística na tendência convergente/divergente global.

A endo/exogeneidade é matematicamente derivada do *corpus*: um termo é endógeno se for frequente no respectivo sub-corpus e exógeno no caso contrário. Ainda não dispomos de elementos suficientes para responder ao terceiro parâmetro.

Para o primeiro parâmetro, temos de calcular a proporção de itens com os traços ‘bi-nacional’ (A<sub>UNI</sub>), ‘endógeno’ (A<sub>ENDO</sub>) e ‘exógeno’ (A<sub>EXO</sub>). Matematicamente, a evolução convergente implica aumento de itens bi-nacionais e de itens exógenos e diminuição de itens endógenos; inversamente, a evolução divergente implica diminuição de itens bi-nacionais e exógenos e aumento de itens endógenos.

O Quadro 3 apresenta os resultados de endo/exogeneidade no *corpus* de futebol: aumento de A<sub>UNI</sub> de 50 para 70, maior do lado do Brasil (B); diminuição de A<sub>ENDO</sub> de 50 para 70, maior do lado de B; maior alteração de A<sub>EXO</sub> em B. Estes resultados condizem com a convergência anteriormente observada (de 50 para 70) e confirmam que o PB muda mais do que o PE.

$A'_{UNIP50,b50}(P50) < A'_{UNIP70,b70}(P70) \cong A'_{UNIP00,b00}(P00)$		
48,47	54,98	58,01
$A'_{EXOP50,b50}(P50) \cong A'_{EXOP70,b70}(P70) \cong A'_{EXOP00,b00}(P00)$		
4,09	4,99	4,24
$A'_{ENDOP50,b50}(P50) > A'_{ENDOP70,b70}(P70) \cong A'_{ENDOP00,b00}(P00)$		
47,44	40,03	37,74
$A'_{UNIB50,p50}(B50) < A'_{UNIB70,p70}(B70) \cong A'_{UNIB00,p00}(B00)$		
38,69	55,82	55,84
$A'_{EXOB50,p50}(B50) \cong A'_{EXOB70,p70}(B70) \cong A'_{EXOB00,p00}(B00)$		
9,53	5,56	5,93
$A'_{ENDOB50,p50}(B50) > A'_{ENDOB70,p70}(B70) \cong A'_{ENDOB00,p00}(B00)$		
51,78	38,62	38,23

**Quadro 3. Evolução dos termos de futebol bi-nacionais, endógenos e exógenos**

Todavia, a medição mais importante é a que consiste em saber o que é que acontece em 70 com os termos que em 50 são endógenos, exógenos e bi-nacionais e o que é que acontece em 2000 relativamente a 1970. Os resultados estão no Quadro 4.

$A'UNIp50,b50(P50) \cong A'UNIp50,b50(P70)$ 48,47                      49,98	$A'UNIp50,p50(B50) < A'UNIp50,p50(B70)$ 38,69                      45,32
$A'EXOp50,b50(P50) \cong A'EXOp50,b50(P70)$ 4,09                        4,86	$A'EXOb50,p50(B50) < A'EXOb50,p50(B70)$ 9,53                        18,54
$A'ENDOp50,b50(P50) \cong A'ENDOp50,b50(P70)$ 47,44                      44,65	$A'ENDOb50,p50(B50) > A'ENDOb50,p50(B70)$ 51,78                      34,50
$A'UNIp70,b70(P70) < A'UNIp70,b70(P00)$ 54,98                      60,10	$A'UNIp70,p70(B70) \cong A'UNIp70,p70(B00)$ 55,82                      55,10
$A'EXOp70,b70(P70) \cong A'EXOp70,b70(P00)$ 4,99                        5,56	$A'EXOb70,p70(B70) \cong A'EXOb70,p70(B00)$ 5,56                        8,00
$A'ENDOp70,b70(P70) > A'ENDOp70,b70(P00)$ 40,03                      33,72	$A'ENDOb70,p70(B70) \cong A'ENDOb70,p70(B00)$ 38,62                      35,48

**Quadro 4. Evolução sequenciada dos termos de futebol bi-nacionais, endógenos e exógenos**

De 50 para 70, o PB muda mais do que o PE, donde resulta uma aproximação do PB ao PE, cujo factor indicaremos adiante. Verifica-se também que o aumento de EXO é quase sempre maior do que a diminuição de ENDO tanto no PB como no PE, mas mais ainda naquele do que neste; isto é, as duas variedades adoptam mais itens exógenos do que abdicam de termos endógenos.

Passemos aos termos de vestuário. Como se pode verificar no Quadro 5, a divergência acima verificada faz-se pela diminuição de termos bi-nacionais e, sobretudo, pelo aumento de termos endógenos (17% no PE e 25% no PB).

$A'_{UNIp50,b50}(P50) > A'_{UNIp70,b70}(P70) > A'_{UNIp00,b00}(P00)$		
75,76	67,10	57,78
$A'_{EXOp50,b50}(P50) \cong A'_{EXOp70,b70}(P70) \cong A'_{EXOp00,b00}(P00)$		
3,31	3,61	4,50
$A'_{ENDOp50,b50}(P50) < A'_{ENDOp70,b70}(P70) < A'_{ENDOp00,b00}(P00)$		
20,93	29,30	37,72
$A'_{UNIB50,p50}(B50) > A'_{UNIB70,p70}(B70) > A'_{UNIB00,p00}(B00)$		
82,20	65,50	54,84
$A'_{EXOb50,p50}(B50) \cong A'_{EXOb70,p70}(B70) \cong A'_{EXOb00,p00}(B00)$		
1,83	4,59	4,74
$A'_{ENDOb50,p50}(B50) < A'_{ENDOb70,p70}(B70) < A'_{ENDOb00,p00}(B00)$		
15,96	29,91	40,42

**Quadro 5. Evolução dos termos de vestuário bi-nacionais, endógenos e exógenos**

E o Quadro 6 mostra três aspectos. Primeiro, a percentagem do vocabulário comum em 50 ou em 70 diminui sempre no período seguinte; o que condiz com a divergência. Segundo, temos agora um fenómeno estranho, sobretudo de 50 para 70: o PE abdica de uma parte dos termos endógenos no período anterior (de 20,93% para 13,69), aumenta a parte dos termos exógenos e, do outro lado, o PB aumenta a parte dos termos exógenos. O estranho é que isto não conduz a uma maior uniformidade com o PB em 70. Todavia, e é este o terceiro ponto, nos anos 70 surgem novos termos endógenos tanto de um lado como do outro: mais 15,61% no PE (29,30% menos 13,69%) e mais 14,7% no PB (29,91% menos 15,21%), o que explica a evolução divergente de ambos os lados. Em 2000 surgem, no PE, mais 4,01% de termos endógenos novos e mais 4,41% de termos que são endógenos nos anos 70; e em PB, mais 9,48 % de termos endógenos novos e mais 1,03% de termos endógenos em 70; o que continua a explicar a evolução divergente de ambos os lados. Esta endogeneidade nova tem a ver com a entrada no mundo da moda, em geral, e do vestuário, em particular, de termos novos (quase sempre estrangeirismos) ou de usos novos de termos já existentes.

$A'_{UNIp50,b50}(P50) > A'_{UNIp50,b50}(P70)$ 75,76                      65,15	$A'_{UNIB50,p50}(B50) > A'_{UNIB50,p50}(B70)$ 82,20                      61,23
$A'_{EXOp50,b50}(P50) < A'_{EXOp50,b50}(P70)$ 3,31                        9,41	$A'_{EXOb50,p50}(B50) < A'_{EXOb50,p50}(B70)$ 1,83                        7,64
$A'_{ENDOp50,b50}(P50) > A'_{ENDOp50,b50}(P70)$ 20,93                      13,69	$A'_{ENDOb50,p50}(B50) \cong A'_{ENDOb50,p50}(B70)$ 15,96                      15,21
$A'_{UNIp70,b70}(P70) > A'_{UNIp70,b70}(P00)$ 67,10                      52,86	$A'_{UNIB70,p70}(B70) > A'_{UNIB70,p70}(B00)$ 65,50                      57,62
$A'_{EXOp70,b70}(P70) \cong A'_{EXOp70,b70}(P00)$ 3,61                        7,41	$A'_{EXOb70,p70}(B70) \cong A'_{EXOb70,p70}(B00)$ 4,59                        7,55
$A'_{ENDOp70,b70}(P70) \cong A'_{ENDOp70,b70}(P00)$ 29,30                      33,71	$A'_{ENDOb70,p70}(B70) \cong A'_{ENDOb70,p70}(B00)$ 29,91                      30,94

**Quadro 6. Evolução sequenciada dos termos de vestuário bi-nacionais, endógenos e exógenos**

Vejamos agora o impacto dos estrangeirismos na tendência evolutiva global. Vamos medir a proporção de termos com o traço (A) ‘anglicismo’, ‘francesismo’ ou ‘estrangeirismo’ (qualquer que seja a proveniência) no perfil onomasiológico de um conceito e, depois, no conjunto dos conceitos analisados, quer do futebol quer do vestuário, nas amostras do PE e PB. É importante notar que neste cálculo a atribuição do traço 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) a fortes adaptações e a decalques.

Os Quadros 7 e 8 apresentam os resultados com as percentagens de anglicismos ( $A'_{Ingl}$ ), francesismos ( $A'_{Fr}$ ) e todos os estrangeirismos ( $A'_{estrangeir}$ ).

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

**Quadro 7. Estrangeirismos no corpus de futebol**



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' Ingl (P50)	3,3%	4,2%	A' Ingl (B50)
A' Ingl (P70)	5,8%	7,6%	A' Ingl (B70)
A' Ingl (P00)	16,9%	17,0%	A' Ingl (B00)
A' estrang (P50)	22,4%	23,8%	A' estrang (B50)
A' estrang (P70)	22,1%	26,7%	A' estrang (B70)
A' estrang (P00)	28,2%	24,9%	A' estrang (B00)

**Quadro 8. Estrangeirismos no *corpus* de vestuário**

No futebol, 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 maior número e frequência de estrangeirismos conservando a sua forma original no PB. Estes resultados contribuem, em parte, para uma distância maior entre as duas variedades nos anos 50 e, conseqüentemente, para a tendência convergente global.

No vestuário, verifica-se uma diminuição de francesismos (mais acentuada no PB) e um claro aumento de anglicismos, por razões sócio-culturais bem conhecidas. É interessante verificar que a influência do francês não é menos importante no PB do que no PE. 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.

Estes resultados confirmam a hipótese de uma maior permeabilidade do PB aos estrangeirismos, quer importando-os directamente, quer adaptando-os. Em relação à adaptação de termos estrangeiros, o Quadro 9 mostra bem a maior tendência do PB à adaptação, em contraste com o PE, que tende para a substituição por termos vernáculos. A percentagem de adaptações de termos ingleses de futebol 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.

A' <sub>Ingl.adapt</sub> (P50)	6,0%	2,8%	A' <sub>Ingl.adapt</sub> (B50)
A' <sub>Ingl.adapt</sub> (P70)	7,9%	16,9%	A' <sub>Ingl.adapt</sub> (B70)
A' <sub>Ingl.adapt</sub> (P00)	8,9%	16,0%	A' <sub>Ingl.adapt</sub> (B00)

### Quadro 9. Adaptações/decalques de anglicismos no *corpus* do futebol

Foram ainda calculadas as proporções de três outros traços ligados aos itens lexicais. Primeiro, há mais mudanças onomasiológicas formais, incluindo arcaísmos, neologismos e termos de gíria, no PB do que no PE. Segundo, a variação onomasiológica formal diminui no futebol (crescente standardização do vocabulário do futebol), em grau mais acentuado no PB, como efeito de popularidade e globalização. Finalmente, a esperada influência do PB no PE, em particular no domínio do futebol, não se confirma claramente, como se pode verificar no Quadro 10. São apresentados 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' <sub>Bras</sub> (P50)	0,8%	2,3%
A' <sub>Bras</sub> (P70)	1,0%	3,4%
A' <sub>Bras</sub> (P00)	1,1%	2,0%

### Quadro 10. Brasileirismos no *corpus* português de futebol

Vamos agora ver até que ponto é que os principais traços ligadas a itens influenciam os valores de uniformidade (U) e, assim, contribuem para a tendência evolutiva global. Para o efeito, utilizamos o método da *análise de regressão*: o valor *p* indica se a relação em causa é ou não significativa,

sendo estatisticamente significativa se  $p$  for inferior a 0,05 (um valor entre 0,05 e 0,10 pode ser tomado como indicando uma tendência); o valor  $r^2$  indica que percentagem dos dados é explicada pela variável em causa. Como se pode verificar pelo Quadro 11, a variável ENDO (termos endógenos) é a que mais se correlaciona com U, já que apresenta sempre resultados estatísticos significativos (e uma média de 99% da evolução de U' pode ser descrita de acordo com a fórmula A'ENDO); as variáveis termos do INGL e todos os ESTR (estrangeirismos) são também significativas no vocabulário do futebol no PE; as outras correlações não passam no teste de  $p$ . Mais especificamente, há uma correlação negativa entre U' e ENDO' e, com excepção do futebol no PE, entre U' e INGL' e ESTR (e também EXO', embora não significativa), isto é, U' diminui quando a proporção destas variáveis aumenta; e há uma correlação positiva, embora não significativa, entre U' e FRAN' no vocabulário do vestuário. Todavia, estes resultados devem ser tomados com cuidado, visto que o número de factores analisados é baixo (em particular, o número de momentos temporais: apenas 3).

Futebol – Portugal			Vestuário - Portugal		
	$p$	$r^2$ (%)		$p$	$r^2$ (%)
U' = -69,43 + 27,33 A'EXO	0,6389	28,87	U' = 139,7 - 19,03 A'EXO	0,2462	85,78
U' = 110,6 - 1,406 A'ENDO	<b>0,0734</b>	<b>98,68</b>	U' = 105,142 - 1,292 A'ENDO	<b>0,0698</b>	<b>98,8</b>
U' = 13,98 + 4,199 A'INGL	<b>0,004</b>	<b>100</b>	U' = 78,991 - 1,352 A'INGL	0,2916	80,45
U' = 14,36 + 4,94 A'INGL>PORT	0,1488	94,63	U' = 22,57 + 3,069 A'FRAN	0,2613	84,08
U' = 4,433 + 2,831 A'ESTR	<b>0,0049</b>	<b>99,99</b>	U' = 165,8 - 4,064 A'ESTR	0,4298	60,94
Futebol – Brasil			Vestuário - Brasil		
	$p$	$r^2$ (%)		$p$	$r^2$ (%)
U' = 74,97 - 3,292 A'EXO	0,1254	96,17	U' = 93,84 - 7,141 A'EXO	0,2355	86,93
U' = 91,41 - 0,9214 A'ENDO	<b>0,0555</b>	<b>99,24</b>	U' = 92,86 - 0,8893 A'ENDO	<b>0,0171</b>	<b>99,93</b>
U' = 198,6 - 8,581 A'INGL	0,2617	84,03	U' = 84,23 - 1,766 A'INGL	0,237	86,77
U' = 41,07 + 0,9099 A'INGL>PORT	0,1079	97,15	U' = 34,68 + 2,197 A'FRAN	0,3808	68,28
U' = 765,8 - 30,77 A'ESTR	0,5594	40,72	U' = 462 - 15,74 A'ESTR	0,6861	22,4

Quadro 11. Correlação entre uniformidade e traços ligados a itens

#### 4. CONCLUSÕES E TRABALHO FUTURO

A investigação realizada permite as seguintes conclusões. Primeiro, os resultados diferentes dos termos de futebol e de vestuário relativamente à questão principal da presente investigação mostram a complexidade do es-

tudo da convergência e divergência entre o PE e o PB. Segundo, a hipótese da divergência confirma-se no campo lexical do vestuário, mas não no do futebol: os termos de vestuário são mais representativos do vocabulário comum e, por isso, os resultados do vestuário estarão provavelmente mais próximos da realidade sociolinguística; a ligeira convergência de 50 para 70 no campo do futebol é provavelmente um efeito da globalização e da standardização do vocabulário do futebol; há muitas diferenças entre o PE e o PB nos dois campos lexicais, sendo a uniformidade actual global de apenas 57%. Terceiro, não parece haver uma orientação de uma variedade em relação à outra: PE e PB divergem um do outro no vestuário; a ligeira aproximação do PB ao PE no futebol resulta, sobretudo, da adaptação de estrangeirismos, que entraram em grande quantidade no PB, nos anos 50, na forma original; a influência do PB sobre o PE, sobretudo no vocabulário do futebol, é menor do que o que se esperava. Quarto, o PB muda mais do que o PE: será um efeito da maior complexidade externa, da maior variação social ou de um atraso de standardização? Provavelmente um pouco de tudo isto. Quinto, confirma-se que a influência estrangeira, de anglicismos e outros estrangeirismos, é maior no PB: o PB importa e adapta mais estrangeirismos do que o PE. Finalmente, 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 importância e a inevitabilidade da sociolinguística cognitiva.

Naturalmente que mais trabalho há a fazer sobre CONDIV entre PE e PB. Como extensões actuais vamos incluir (i) palavras de outros campos lexicais, como o da saúde, e (ii) palavras funcionais, particularmente preposições. Como extensão futura, queremos incluir o domínio da gramática, analisando variáveis não-lexicais (morfológicas e sintácticas) e as correlações entre estas e as variáveis lexicais. Tendo em conta a distinção entre léxico (mais consciência) e gramática (menos consciência), a hipótese é a de maior divergência: as palavras funcionais e as construções gramaticais divergirão mais do que as palavras de conteúdo.

## REFERÊNCIAS BIBLIOGRÁFICAS

GEERAERTS, Dirk (2005). Lectal variation and empirical data in Cognitive Linguistics. In: Ruiz de Mendoza, F. J. & Peña Cervel, S. (eds.), *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interactions*. Berlin/New York: Mouton de Gruyter, 163-189.

GEERAERTS, Dirk, GRONDELAERS, Stefan & SPEELMAN, Dirk (1999). *Convergentie en divergentie in de Nederlandse woordenschat*. Amsterdam: Meertens Instituut.

KRISTIANSEN, Gitte & DIRVEN, René (eds.) (2007). *Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter.

SILVA, Augusto Soares da (2006a). O léxico do futebol no Português Europeu e no Português Brasileiro: convergência ou divergência? *Diacrítica. Ciências da Linguagem* 20-1. 167-196.

SILVA, Augusto Soares da (2006b). Sociolinguística cognitiva e o estudo da convergência/divergência entre o Português Europeu e o Português Brasileiro. *Veredas* 10. *Revista de Estudos Lingüísticos*. <[www.revistaveredas.ufjf.br](http://www.revistaveredas.ufjf.br)>

SILVA, Augusto Soares da (2007). Integrando a variação social e métodos quantitativos na investigação da linguagem e cognição. Para uma sociolinguística cognitiva do Português Europeu e Brasileiro. Plenária apresentada na *IV Conferência Lingüística e Cognição*, Universidade Federal de Minas Gerais, 9-11 Agosto 2007 (a publicar na revista *Estudos da Linguagem*, da UFMG).

SILVA, Augusto Soares da & DUARTE, Marlene (2005). O léxico do vestuário no Português Europeu e no Português Brasileiro: convergência ou divergência? *Revista Portuguesa de Humanidades* 9, 117-136.