

O corpus CONDIV e o estudo da convergência e divergência entre variedades do português

Augusto Soares da Silva

Universidade Católica Portuguesa – Braga

assilva@braga.ucp.pt

Colóquio Linguateca – 10 Anos

Curia, 11 Setembro 2008

Objetivos

- apresentar o CONDIV como instrumento para o estudo da convergência e divergência entre PE e PB
- situar o CONDIV num programa de sociolinguística cognitiva e variação do português
- promover o desenvolvimento do CONDIV

Sumário

1. CONDIVport: objetivos e estrutura
2. Contexto teórico e metodologia: o CONDIV na articulação entre cognitivo, social e empírico
3. Convergência ou divergência entre PE e PB?
Resultados da investigação sociolexicológica
4. Extensões lexicais e gramaticais

O corpus CONDIVport: para quê?

1. convergência ou divergência entre PE e PB nos últimos 50 anos?
→ análise **diacrónica externa**
2. como é que factores linguísticos internos influenciam a convergência/divergência global?
→ análise **diacrónica interna**
3. estratificação actual do PE e do PB (distância entre estratos): idêntica ou diferente?
→ análise **sincrónica**

O corpus CONDIVport: o quê?

- variáveis
 - geográfica: Portugal vs. Brasil
 - diacrónica: 1950 - 1970 - 1990/2000
 - estilística: jornais, revistas de qualidade > populares
> Internet *off-line* > Internet *on-line* (chats)
- domínios: futebol, moda/vestuário, saúde ...
- textos extraídos (5 milhões + 15 milhões) de
 - jornais desporto (8), revistas moda (30) e saúde (20)
 - linguagem Internet: *chats* (IRC)
 - lojas de vestuário: etiquetas
- acesso: parte em Linguateca AC/DC

Quadro teórico

Linguística Cognitiva. Porquê?

- LC é:
 - modelo orientado para o significado
 - modelo experiencialista
 - modelo baseado-no-uso
 - modelo recontextualizador
- cada uma destas características implica:
 - ir do cognitivo ao social, mas há problemas: é preciso reinterpretar a cognição como *cognição social*
 - opção pelos métodos empíricos, mas há resistências

Sociolinguística Cognitiva

- perspectiva cognitiva dos fenómenos variacionais: aplicação de modelos descritivos (TP, TMC, etc.)
- exploração da relação entre cognição e variação lectal
- exploração da *cognição social*: elucidação da interacção dialéctica entre o nível individual cognitivo e o nível social das normas colectivas
- desenvolvimento de métodos quantitativos de corpus e métodos de análise multivariacional da confluência de factores conceptuais, discursivos e variacionais

Variação lexical

1. variação semasiológica

diferentes sentidos/referentes de um item lexical, i.é, polissemia e prototipicidade (cf. Silva 2006)

2. variação onomasiológica conceptual

diferentes categorias conceptuais para nomear um referente, por ex. escolhas taxionómicas: *calças/jeans*

3. variação onomasiológica formal

diferentes termos para nomear uma mesma categoria conceptual, i.é, sinónimos denotacionais: *atacante/avançado, casaco/blazer*

→ **variação contextual/lectal**: adicional a 1 e 2; intrínseca a 3

Estudar convergência e divergência

O quê?

variação onomasiológica formal – escolhas entre
sinónimos denotacionais

Porquê?

os sinónimos denotacionais revelam diferenças
sociolinguísticas

Como?

Lexicologia Quantitativa e Sociolectometria:

- medidas de uniformidade **U**
- medidas do impacto de traços **A**

Medida U

uniformidade linguística baseada em perfis

- **perfil onomasiológico**: conjunto de sinónimos usados para designar um conceito, diferenciados pela sua frequência relativa
- **uniformidade**: medida da correspondência entre dois conjuntos de dados, definidos em termos de perfis onomasiológicos
 - medida **U**: sem ter em conta a frequência
 - medida **U'**: tendo em conta a frequência

formalmente:

medida **U**

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

Uniformidade U para o conceito Z entre duas amostras Y1 e Y2 =
soma dos mínimos
das frequências relativas F
do item lexical x
nos perfis onomasiológicos para Z
em Y1 e Y2

medida U'

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

Uniformidade U' para o conjunto de conceitos Z

entre duas amostras Y_1 e $Y_2 =$

soma dos valores- U dos Z s

ponderada pela frequência relativa G
de Z dentro do conjunto total de Z s

FUTEBOL	P50		B50		U	U'
AVANÇADO						
atacante	101	8,8	119	36,6		
avançado	820	71,6	3	0,9		
avante	0	0,0	159	48,9		
dianteiro	220	19,2	22	6,8		
"forward"	1	0,1	17	5,2		
ponta-de-lança	3	0,3	5	1,5		
					16,9	0,6

Medida A

- medida **A**: proporção A de todos os itens x com o traço K no perfil onomasiológico do conceito Z no subcorpus Y = soma das frequências relativas de x's ponderada pelo valor de pertença W

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

- medida **A'** : proporção A' de todos os itens x com o traço K no subcorpus Y = soma de todas as medidas-A, ponderada por G, i.é, a frequência relativa do conceito Z em Y

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

Convergência ou divergência
entre PE e PB?

alguns resultados
da investigação sociolexicológica
(futebol e vestuário)

Futebol

- 21 perfis onomasiológicos (183 termos)
- 2.7 milhões ocorrências (de 8 jornais de desporto)
- 90.202 observações

perfis de futebol

AVANÇADO: atacante, avançado, avante, dianteiro, “forward”, ponta-de-lança.

DEFESA: “(full-)back”, beque, bequeira, defensor, defesa, lateral, líbero, zagueiro.

MÉDIO: alfe, central, centro-campista, centro-médio, “half”, interior, médio, meia, meio-campista, meio-campo, “midfield”, trinco, volante.

EXTREMO: ala, extremo, ponta, ponteiro.

GUARDA-REDES: arqueiro, “goal-keeper”, goleiro, golquíper, guarda-meta, guarda-rede, guarda-redes, guarda-vala, guarda-valas, guardião, “keeper”, porteiro, quíper, vigia.

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”.

BOLA: balão, bola, couro(inho), esfera, esférico, pelota.

BALIZA: arco, baliza, cidadela, “goal”, gol, malhas, marco, meta, rede, redes, vala.

GOLO: bola, “goal”, gol, golo, ponto, tento.

FINTA: corte, “drible(ing)”, engano, “feint”, finta, firula, ginga, lesa, manobra enganadora, simulação.

JOGADA: jogada, lance.

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

FORA-DE-JOGO: adiantamento, banheira, deslocação, fora-de-jogo, impedimento, “offside”, posição irregular.

GRANDE PENALIDADE: castigo máximo, castigo-mor, falta máxima, grande penalidade, penalidade, penalidade máxima, pênalti, penálti, “penalty”.

PONTAPÉ: chute, chuto, “kick(-off)”, panázio, pelotada, pontapé, quique, “shoot”, tiro.

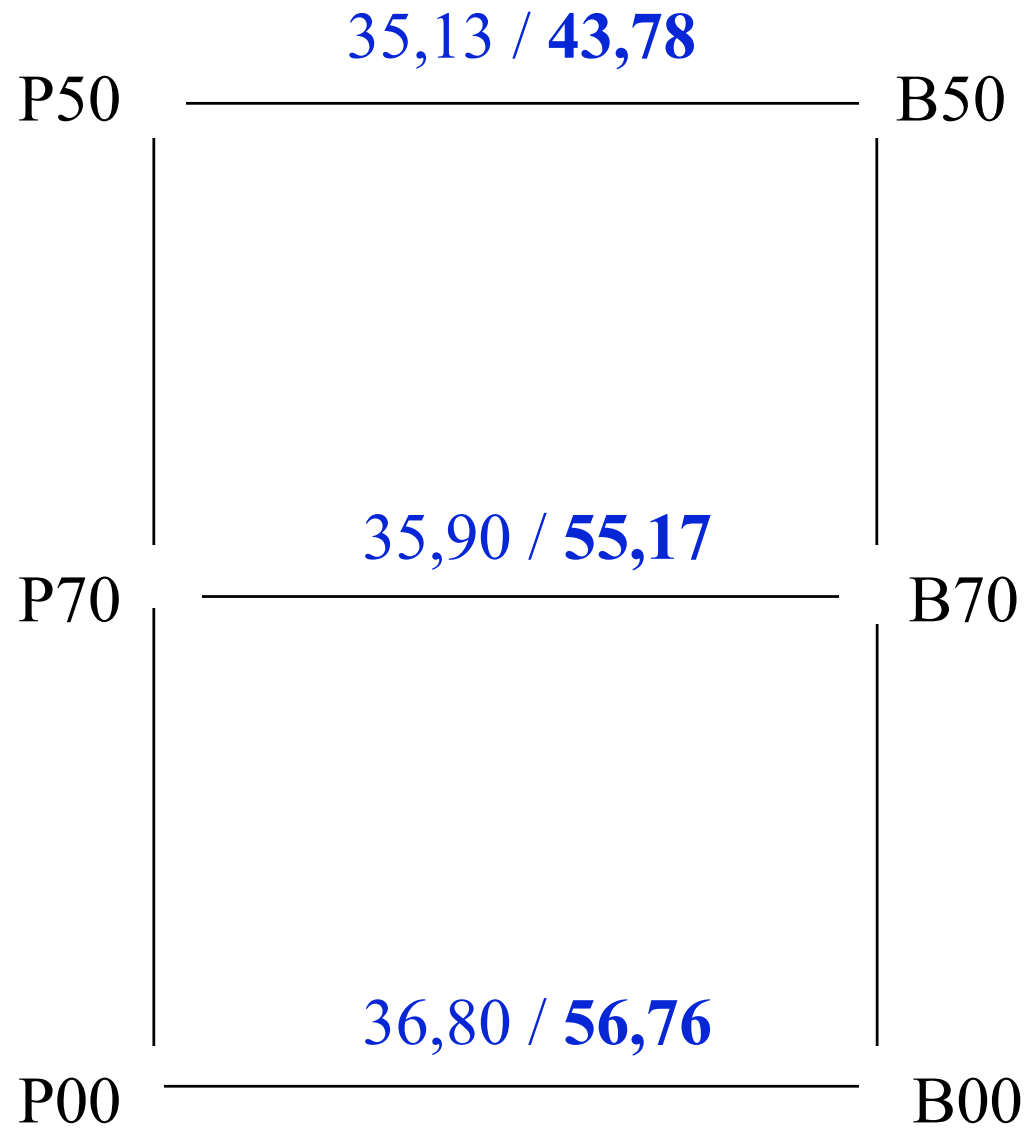
PONTAPÉ DE CANTO: canto, chute de canto, “corner”, córner, escanteio, esquinado, pontapé de canto, tiro de canto.

PONTAPÉ LIVRE: chute (in)direto, falta, “free(-kick)”, livre (directo, indirecto), pontapé livre, tiro dire(c)to, tiro livre (direto, indirecto)

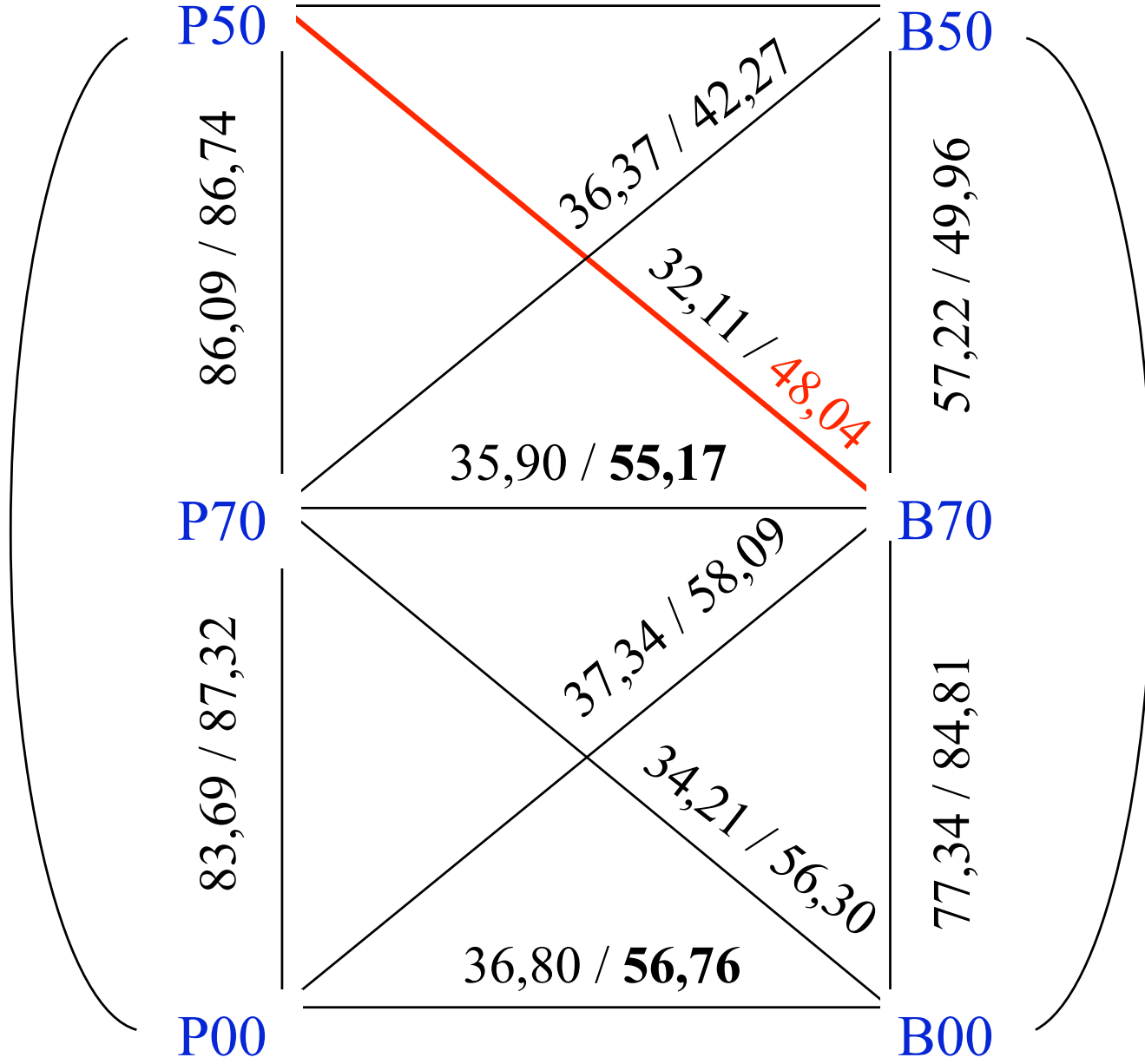
EQUIPA: conjunto, formação, eleven, equipa/e, escrete, esquadra/ão, grupo, “match”, onze, onzena, plantel, quadro, “team”, time, turma.

TREINADOR: mister, professor, técnico, treinador.

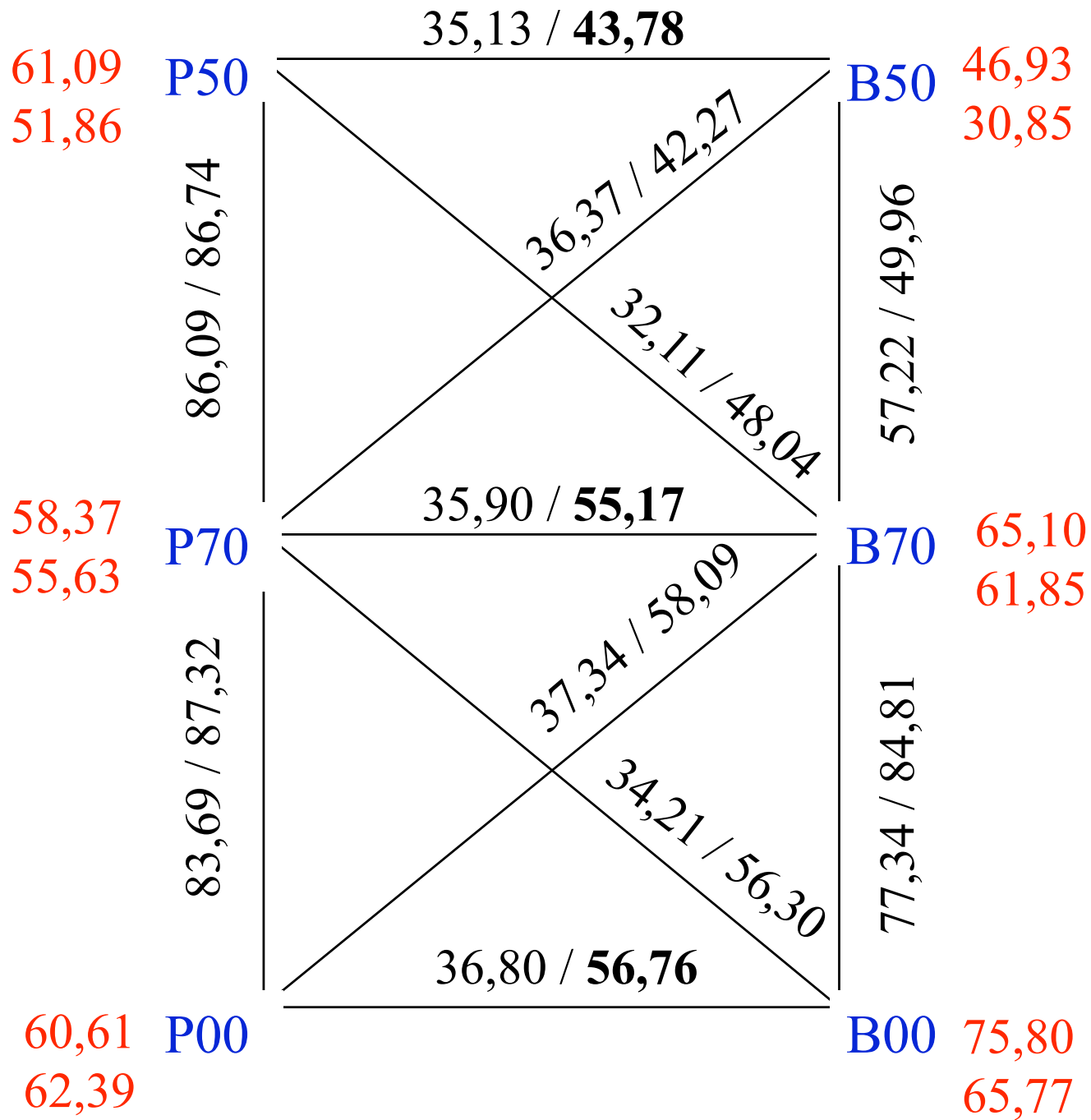
JOGO: batalha, choque, combate, competição, confronto, desafio, disputa, duelo, embate, encontro, jogo, justa, luta, “match”, partida, peleja, prélio, prova, pugna.



76,33 / 78,52



44,89 / 43,24



Endogeneidade/exogeneidade

- aumento de A'_{UNI} dos anos 50 para 70, mais forte em PB
- diminuição de A'_{ENDO} dos anos 50 para 70, mais forte em PB
- grande diminuição de A'_{EXO} em PB

Estrangeirismos

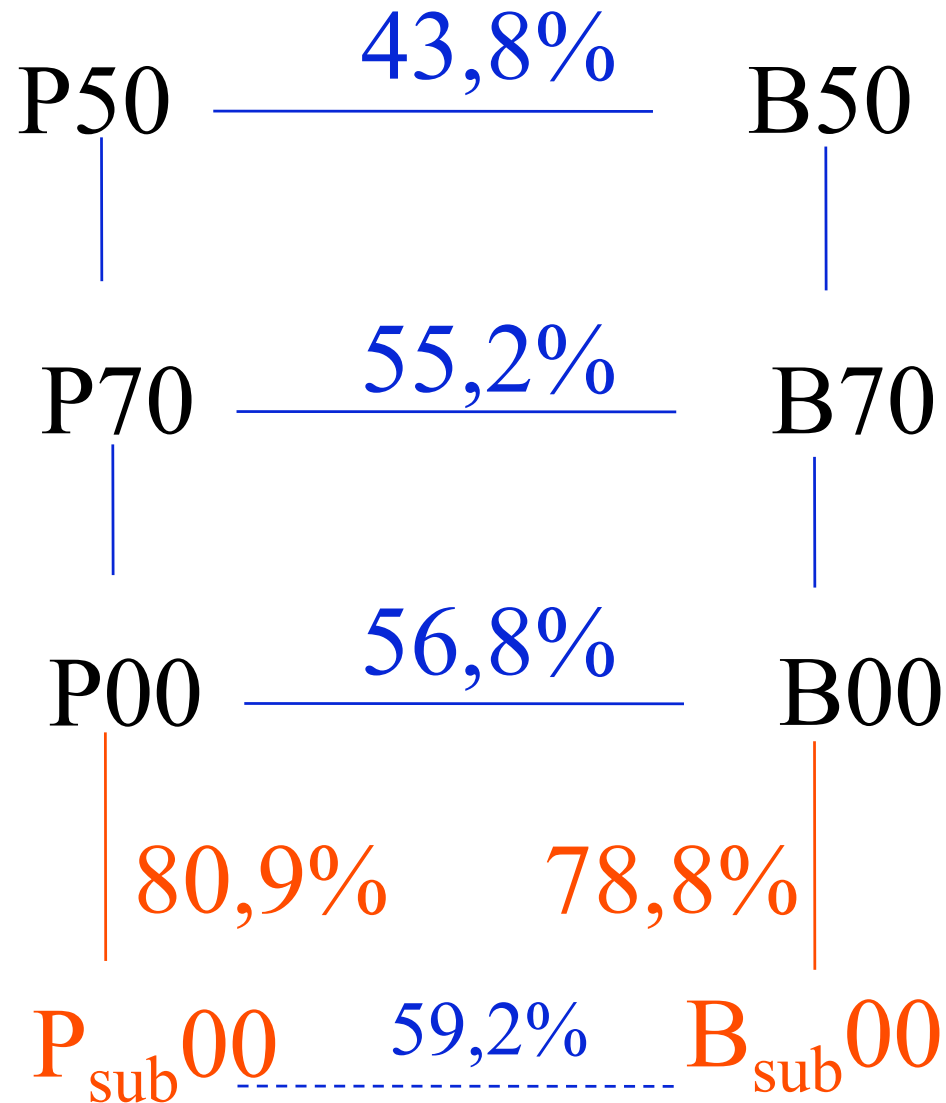
A'_{Ingl} (P50)	7,1%	18,0%	A'_{Ingl} (B50)
A'_{Ingl} (P70)	9,8%	17,1%	A'_{Ingl} (B70)
A'_{Ingl} (P00)	10,2%	16,2%	A'_{Ingl} (B00)

A'_{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)

Brasileirismos no corpus de futebol

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

Estratificação: jornais 2000 vs. chats (143.946 observações)



Vestuário

- 22 perfis onomasiológicos (264 termos)
- 1.2 milhões ocorrências (de 28 revistas de moda)
- 12.451 observações

perfis de vestuário

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

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

JAQUETA M/F: casaca, casaco curto, jaleca, jaqueta, “jaquette”, jaquetinha, véstia

BLUSÃO M/F: blazer, blêizer, blusão, “bluson”, camurça, camurcine, camisa esporte, casaco de pele, ganga, etc., colete, parka

FATO M: beca, completo, costume, fato, terno

CONJUNTO F: “complet”, completo, conjunto, costume, duas-peças, “ensemble”, fatinho, fato, saia-casaco, “tailleur”, “toilette”, toilette, vestido-casaco

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”

CAMISA M: blusão, camisa, camisa de gravata, camisa de manga curta, camisa desportiva, camisa esporte(iva), camisa social, camiseta, camisete, “camisette”, camisinha

BLUSA F: “blouse”, blusa, blusinha, “bustier”, camisa, camisa-body, camiseiro(inho), camiseta/e, (blusa) “chémisier”, (blusa) chemisiê

T-SHIRT M/F: camisa, camiseta/e, “camisette”, camisola, “singlet”, “tee-shirt”, “t-shirt”

CALÇAS M/F: calça, calças, pantalone

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

SAIA F: kilt, maxi (máxi), maxissaia, micro-mini, micro-saia, míni (mini), mini-saia, minissaia, pareô, saia, saia-calça, saia-calção, saião, sainha, saiote

CASACO DE MALHA M/F: cardigã, “cardigan”, casaco/casaquinho de malha (de lã, de tricot), “gilet”, japonsa, malha, “twin-set”

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)”

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 CURTO F: bolero, carmona, casa(i)b(v)equa, casaco curto, casaquilha, colete, colete camiseiro, corpete, corpinho, garibáldi, “gilet”, manguito, mini, minicasaco, roupinha, “shortie”, vasquinha

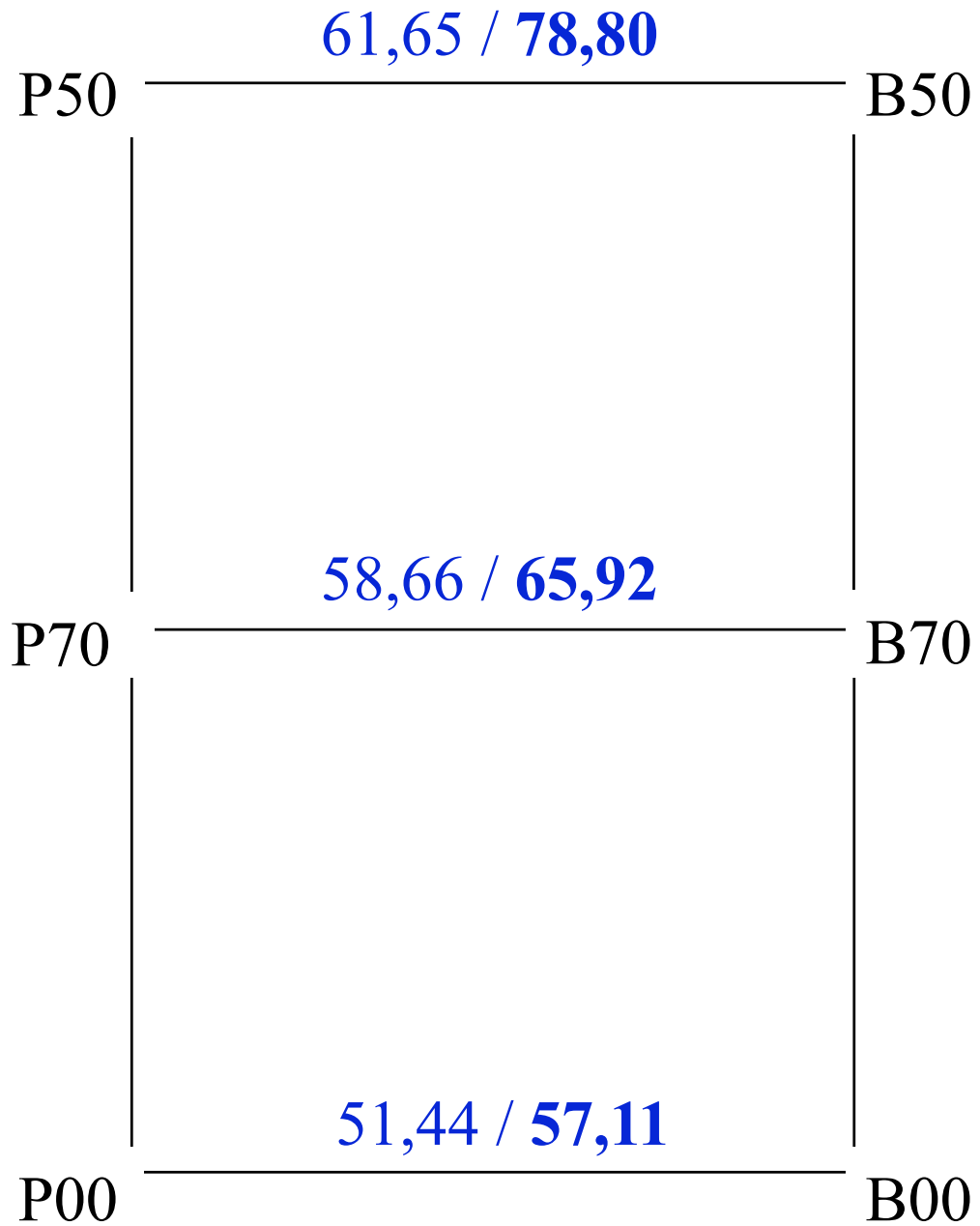
CASACO CURTO M: casaco curto, colete, espartilho, gibão, “gilet”, jaleca, jaleco, jaqueta, véstia

CASACO DE CERIMÓNIA M/F: “black-tie”, casaca, casaco cerimónia, fraque, “manteau”, mantô, paletó, “paletot”, “pelerine”, “smo(c)king”, sobrecasaca, “tuxedo”

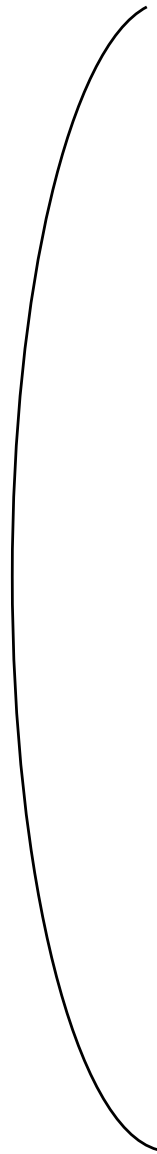
JEANS M/F: calça(s) de ganga, calça(s) em denim / em jeans, ganga, jeans

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)”



60,58 / 65,27



P00

68,25 / 74,40

P70

59,92 / 68,67

P50

51,44 / **57,11**

44,41 / 53,99

58,66 / **65,92**

53,99 / 67,50

61,65 / **78,80**

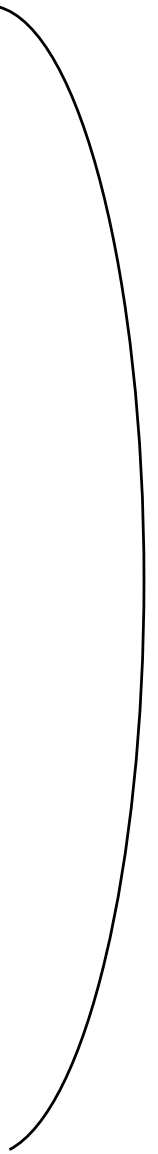
B00

67,70 / 74,74

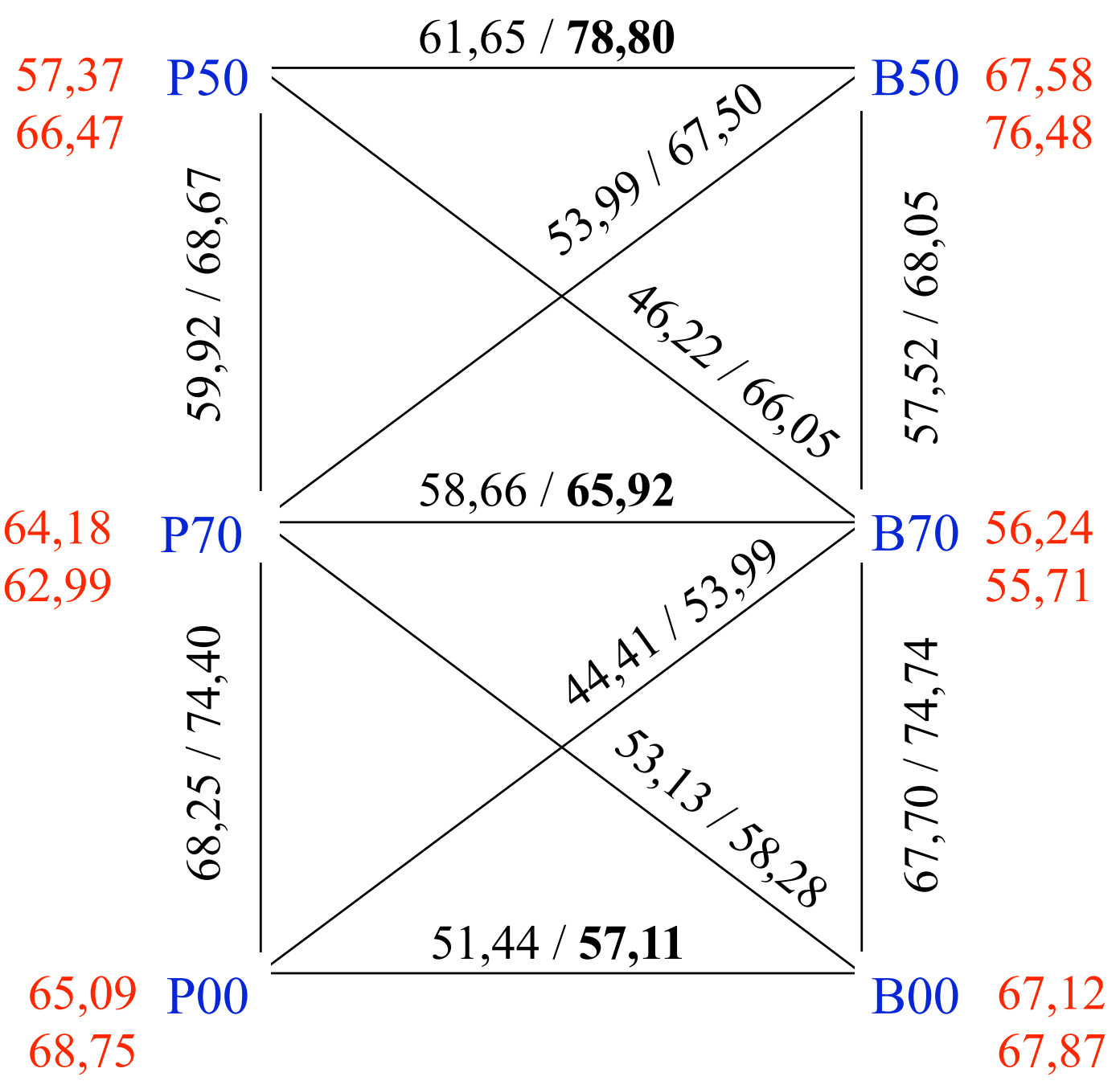
B70

57,52 / 68,05

B50



60,44 / 65,38



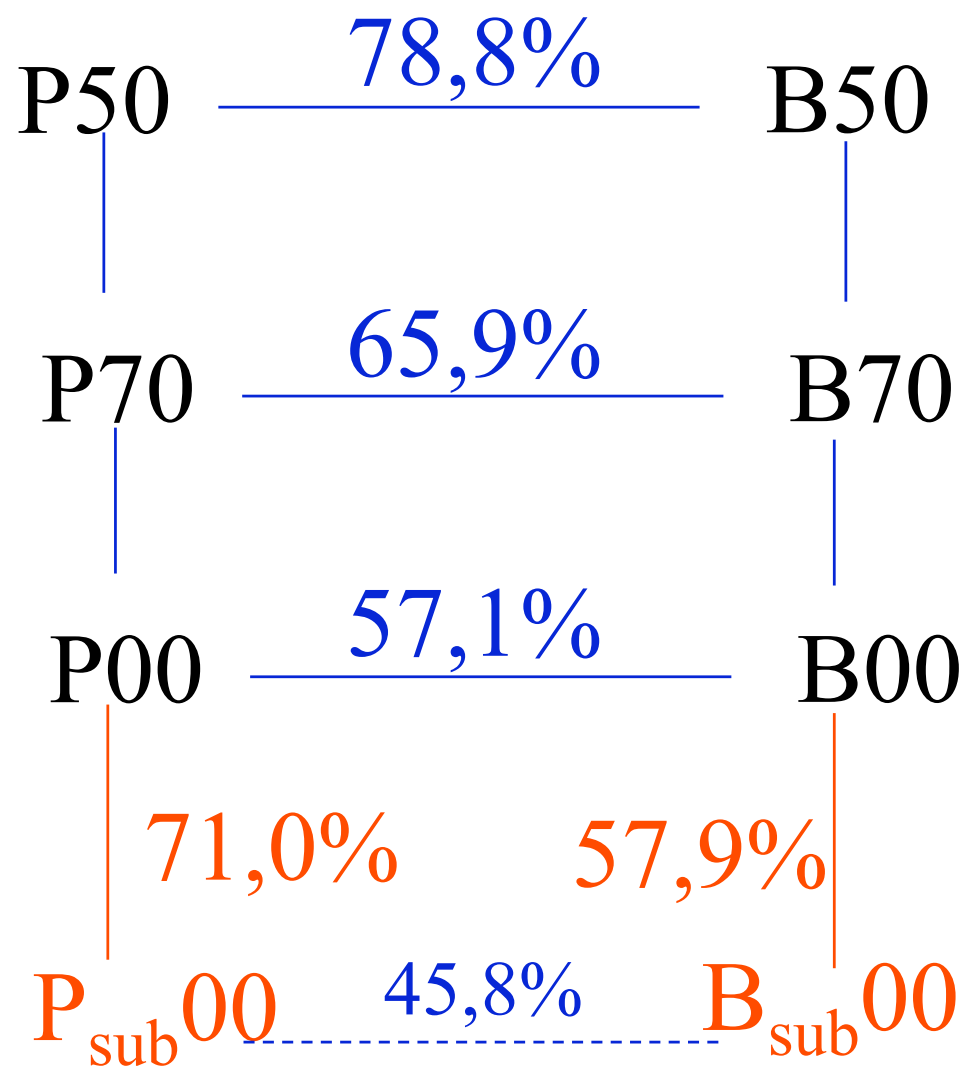
Endogeneidade/exogeneidade

- diminuição de A'UNI e forte aumento de A'ENDO em PE e em PB
→ divergência de ambos os lados
- PB parece mudar mais do que PE, sobretudo de 50 para 70

Estrangeirismos

$A'_{Fr} (P50)$	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%	$A'_{Ing} (B00)$
$A'_{estrang} (P50)$	22,4%	23,8%	$A'_{estrang} (B50)$
$A'_{estrang} (P70)$	22,1%	26,7%	$A'_{estrang} (B70)$
$A'_{estrang} (P00)$	28,2%	24,9%	$A'_{estrang} (B00)$

Estratificação: revistas 2000 vs. etiquetas lojas (3.240 observações)



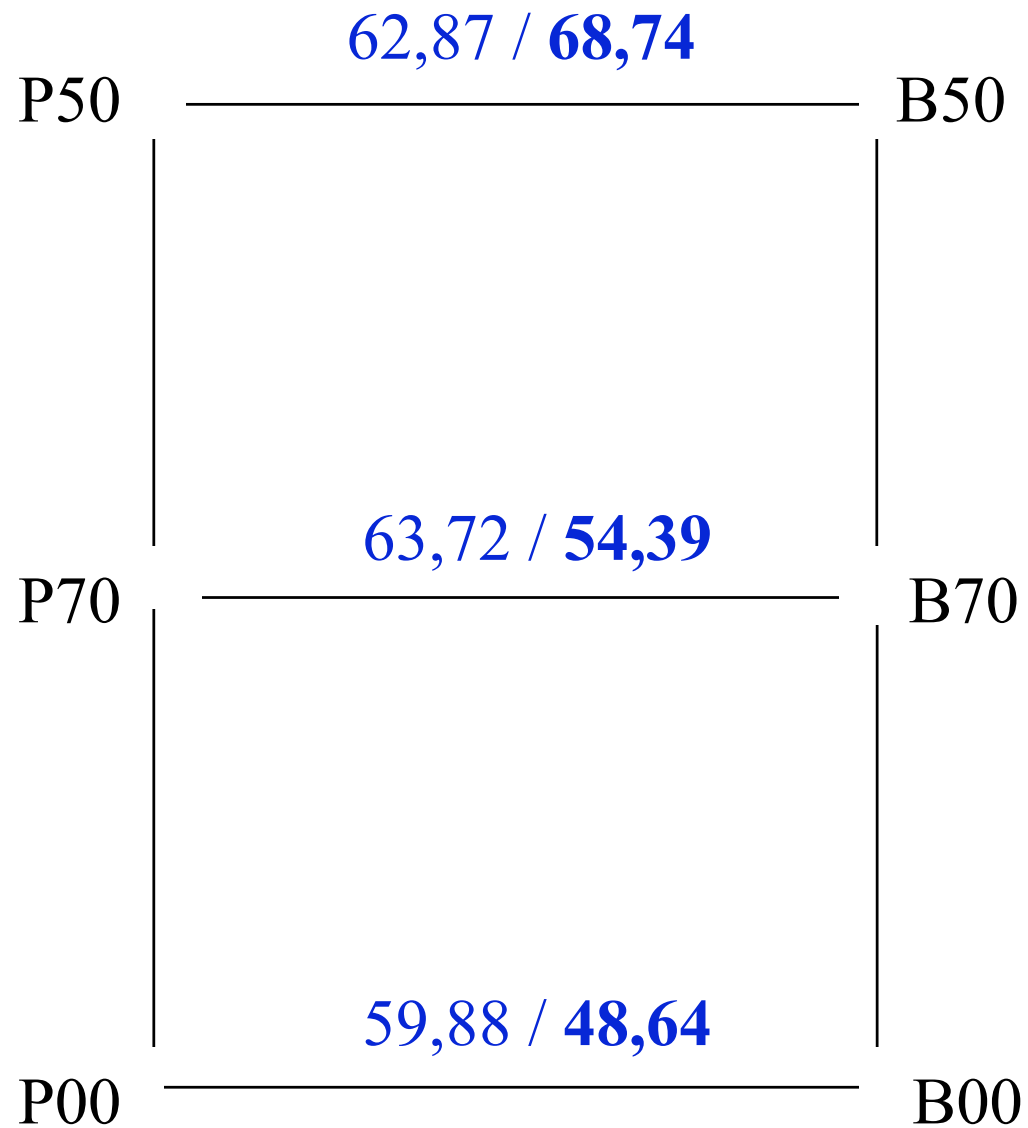
Extensões

lexicais e gramaticais

(preposições e construções)

Palavras funcionais: 10 perfis

- FALAR: *de/sobre/acerca de/em*
- N_{mental}: *acerca de/sobre*
- INTERESSE: *em/por/em torno de*
- PRECISAR, NECESSITAR: *de/∅*
- ANSIOSO: *de/para/por*
- APRESSAR-SE: *a/em/para/por*
- CONVENCER, ESQUECER-SE, LEMBRAR-SE, RECORDAR-SE, ADMIRAR-SE, DUVIDAR, AVISAR: *de que/que*
- CONVENCIDO: *de que/que*
- GOSTAR, FALAR, PRECISAR, NECESSITAR: *relativa/relativa cortadora*
- ESQUECER, LEMBRAR DE: *reflexo/não-reflexo*

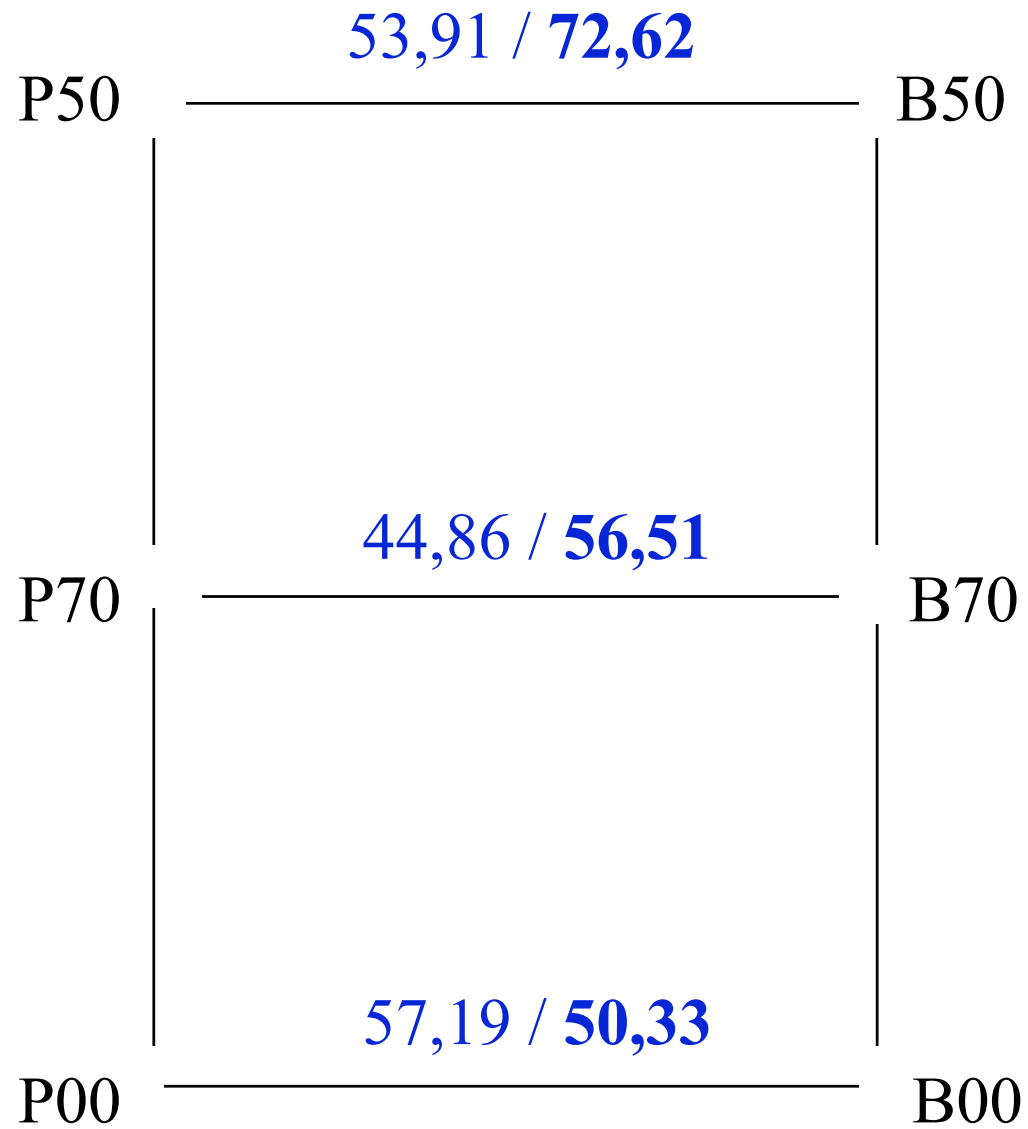


CONDIV-desporto: 1.730 observações

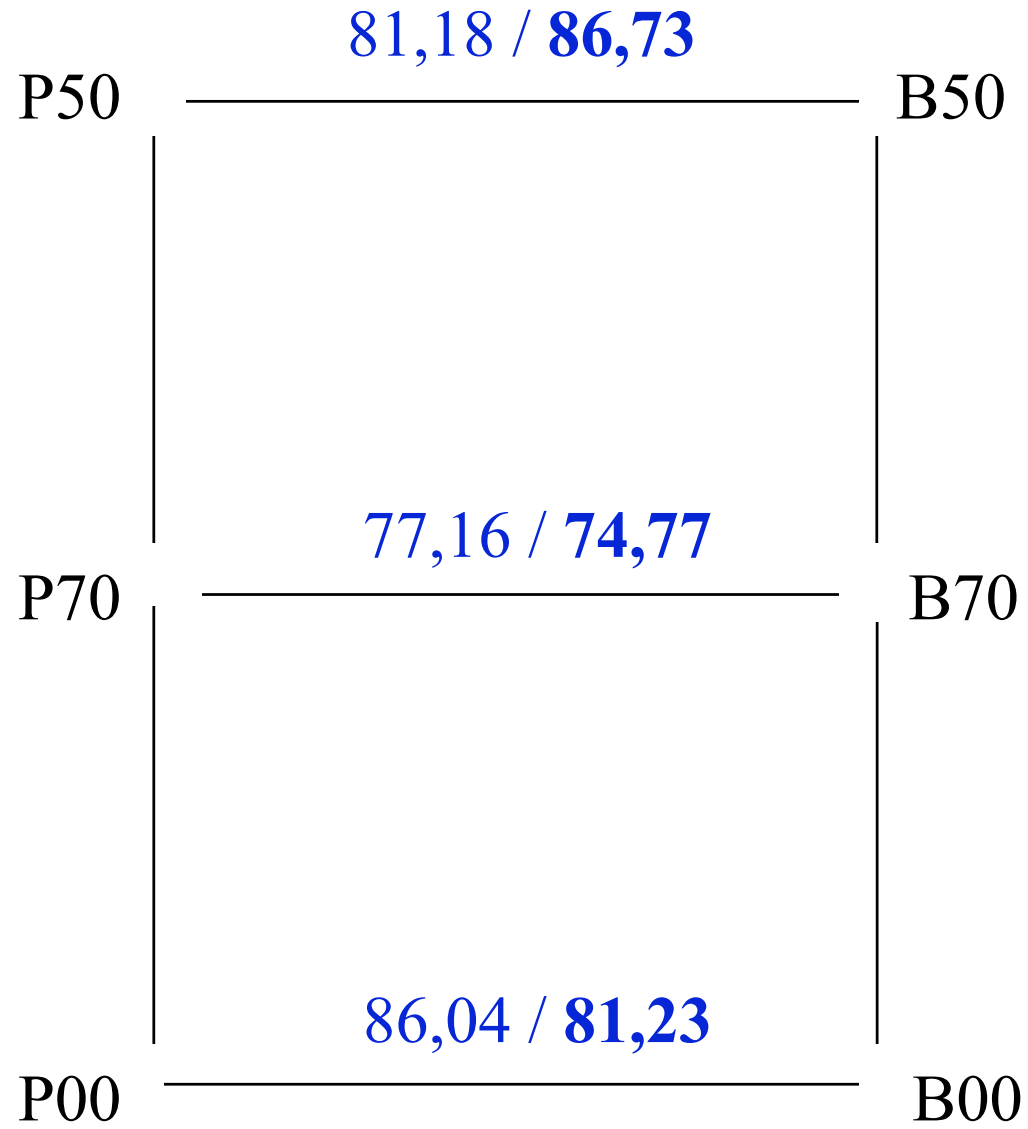
Construções sintáticas: 3 perfis

através do analisador PALAVRAS

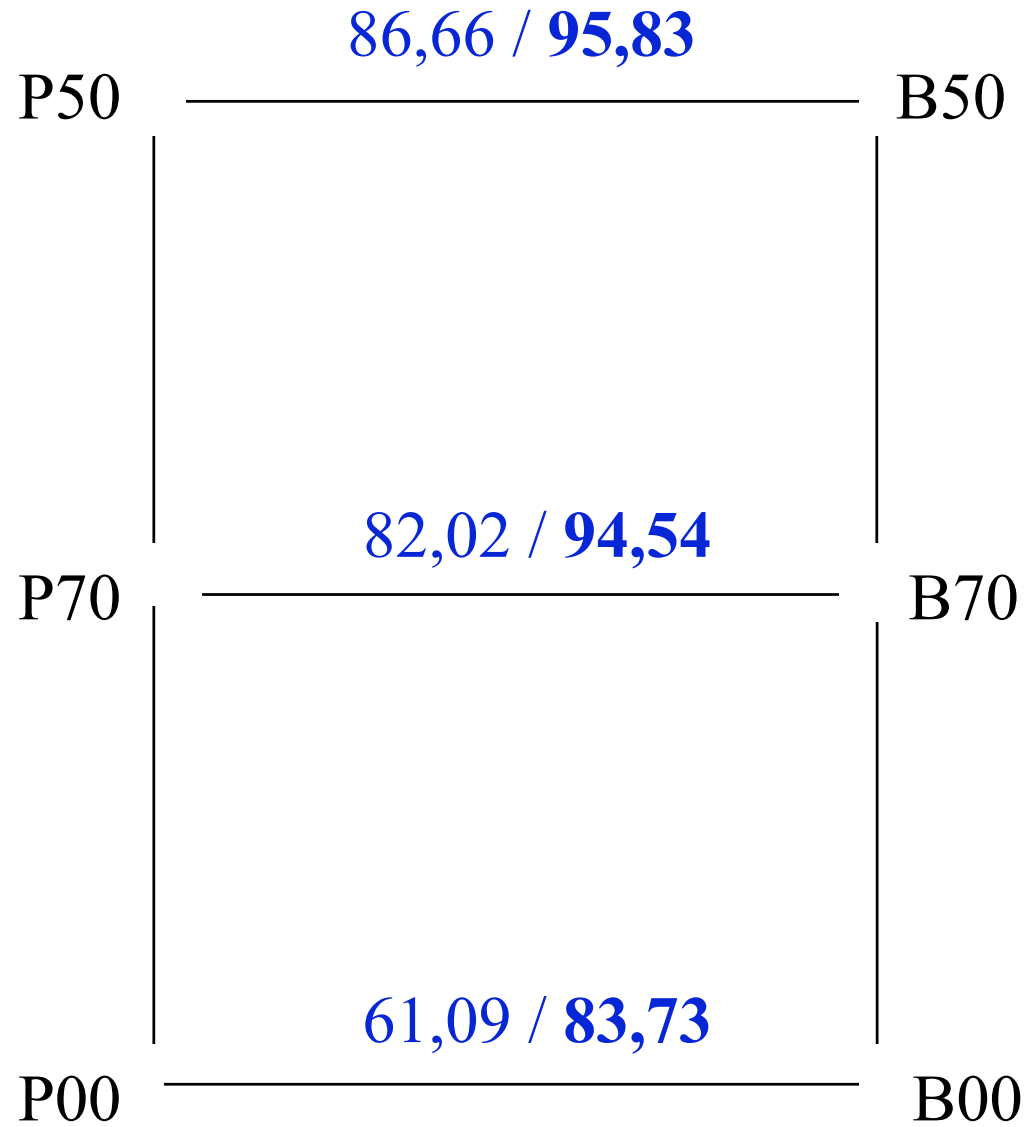
- *deixar, fazer, mandar + ver, ouvir, sentir*
 - VSV CONDIV-desporto
 - VOV 1.385 observações
 - VV
- *deixar, fazer, mandar + ver, ouvir, sentir*
 - INFIN CONDIV-desporto
 - FIN 2.367 observações
- *verdadeiro, falso, bonito, lindo, recente*
 - A+N CONDIV-desporto
 - N+A 763 observações



VSV / VOV / VV



INFIN / FIN



A+N / N+A

Conclusões

1. a importância do CONDIV para o estudo da convergência/divergência entre PE e PB
2. a importância do CONDIV para o estudo da relação entre cognição e variação lectal
3. divergência no vestuário e nas palavras funcionais e construções sintáticas: do léxico à gramática – maior divergência?
4. é preciso aumentar o CONDIV: desafio à Linguateca!
 - mais textos 50-70-2000
 - mais variáveis estilísticas