

PORTUGUÊS EUROPEU E PORTUGUÊS BRASILEIRO: ALGUNS CONTRASTES

Luzia Helena Wittmann e Maria de Jesus Pereira

Grupo de Linguagem Natural - INESC

Setembro 1994

RESUMO Esta comunicação pretende apresentar alguns dos resultados obtidos na primeira fase de uma investigação em curso sobre diferenças entre o Português Europeu e o Português Brasileiro nos níveis lexical, morfo-sintático e ortográfico. Inclui um estudo quantitativo baseado em dois *corpora* paralelos, e elaborado com recurso a ferramentas computacionais.

Pode parecer estranho que, num momento de unificação ortográfica e da criação da Comunidade dos Países de Língua Portuguesa, nos preocupemos em destacar precisamente as diferenças. Sabemos, entretanto, que nenhuma Comunidade poderia medrar sem o devido respeito pela diversidade cultural dos seus membros, bem como sabemos que para que haja respeito é necessário haver consciência, implicando esta, por sua vez, o conhecimento do outro.

A discussão sobre as diferenças e similitudes entre o Português Europeu e o Português Brasileiro não é recente tanto do ponto de vista científico, quanto do político. Há, na verdade, duas histórias dessa polémica que se entrecruzam: a do uso da unidade e da diversidade da língua como instrumento político/ideológico, principalmente a partir do romantismo brasileiro de José de Alencar; e a da evolução dos estudos científicos na área. Mas o que parece haver, sobretudo, é

uma tendência para a confusão e distorção involuntária dos fatos uma vez que, sendo fatos linguísticos, não só são propriedade de cada falante, como passam pela relação afetiva que cada falante tem com sua língua.

Não é nosso objetivo retomar tal polêmica aqui, embora ela se mantenha subjacente a qualquer abordagem do tema. Fomos colocados diante de uma pergunta muito concreta: na hipótese de desenvolvimento de um sistema de tradução automática do Inglês para o Português, vocacionado essencialmente para textos técnicos, é possível abranger as variantes europeia e brasileira da língua, partindo de uma base comum ou será preferível optar por dois sistemas separados?

Foi para responder a essa pergunta que principiámos o estudo, uma vez que não podíamos sequer adiantar um dado aproximado sobre a profundidade das diferenças. Na verdade emerge do senso comum a idéia de que entre o Português Europeu e o Português Brasileiro falado e coloquial há divergências significativas, mas que, no registro escrito padrão (jornalístico por exemplo), nem por isso. Assim, dedicamos três meses a um estudo preliminar que consistiu, numa primeira fase, num levantamento de todos os contrastes que pudemos encontrar nos níveis morfo-sintático e ortográfico. E, numa segunda fase, num levantamento quantitativo dos contrastes *em corpora*, agora incluindo os contrastes lexicais. Pudemos assim chegar a uma primeira idéia aproximada da profundidade das diferenças reais entre as duas variantes.

Toda a investigação, tanto a fase preliminar que estamos descrevendo quanto a sua continuação, está sendo feita em função do objetivo final, ou seja, um sistema de tradução automática, sistema esse direcionado para traduções técnicas. Era importante, portanto, que o estudo contemplasse apenas a língua padrão de cada variante, de modo a excluir diferenças que pertencem à linguagem

oral, popular ou mesmo à linguagem demasiado formal. Os resultados que obtivemos tendem, assim, para um mínimo de contrastes, incluindo apenas aqueles que consideramos incontornáveis. Não sendo pacífico o conceito de língua padrão, consideramos, para a linguagem não técnica, a linguagem corrente, como um registro que se poderia definir por oposição ao registro literário e ao registro popular.

Começamos por recolher todos os contrastes que já haviam de alguma forma sido captados e publicados, tanto por linguistas especializados no tema, quanto por autores de gramáticas tradicionais. Os linguistas especializados resumem-se, com poucas exceções, ao grupo da Universidade de Campinas. Este grupo investiga em profundidade e diacronicamente a especificidade do Português Brasileiro, nos aspectos que o distanciam da variante europeia. Tendo objetivos mais abrangentes, esses estudos obviamente incluem todos os registros da variante brasileira. A partir das conclusões advindas de várias análises rigorosas, defende uma adaptação da atual gramática em vigor no Brasil à realidade lingüística brasileira (cf. Roberts e Kato, orgs., 1993). Entre as poucas exceções, ao menos de que temos notícia, destaca-se o trabalho de Teresa Biderman (Biderman, s/d) que elaborou um vocabulário fundamental de Português Brasileiro. Há também o trabalho diacrônico de Clóvis Barleta de Moraes, sobre a especificidade do Português Brasileiro (Moraes, s/d).

Entre as gramáticas que dão especial atenção ao problema, destaca-se sem dúvida a de Celso Cunha e Lindley Cintra (Cunha, e Cintra, 1987), cuja coragem em assumir as diferenças abriu portas, mesmo se a intenção não era mostrar as diferenças, mas sim as parecenças. O único senão é o fato de considerar, sem discriminação, fenômenos de linguagens particularmente marcadas, como o oral popular, com outros que se manifestam de forma generalizada e constante. Algumas gramáticas dirigidas a estudantes e utentes estrangeiros também se

mostram preocupadas com os contrastes entre as duas variantes. É o caso, por exemplo, da gramática de Pilar Cuesta (Cuesta e Luz, 1971) para espanhóis, a de Paul Teyssier (Teyssier, 1989) para os franceses e a de Helena Abreu e Rita Benamor (Abreu e Benamor, 1994) para os italianos. Esta última, que ainda está no prelo, é a mais atualizada e cuidadosa em relação às diferenças gramaticais entre as duas variantes, uma vez que inclui um livro de exercícios com dupla versão: portuguesa e brasileira

Não deixa de ser sintomático que o próprio mercado editorial de material didático de português para estrangeiros comece a exigir tal diferenciação. Os editores, pelo menos os da Zanichelli (eds. da gramática de Abreu e Benamor) chegaram à conclusão de que existe um público para o Português Europeu e um outro para o Português Brasileiro e tentam abranger os dois. Por outro lado sabemos que as publicações de livros brasileiros em Portugal e vice-versa sofrem uma conversão às vezes só ortográfica (como as edições da obra de Jorge Amado pela Europa-América), outras vezes também lexical e sintática (é o caso dos romances de Paulo Coelho, pela Pergaminho)

Com as ferramentas computacionais a questão torna-se ainda mais premente. Mesmo depois de o acordo ortográfico ser implantado, um corretor ortográfico de Português Europeu será pouco funcional se for usado sem nenhuma adaptação para o Português Brasileiro e vice-versa. Da mesma maneira, um tradutor automático para PE não pode ser usado também para PB sem a devida adaptação.

Feito esse levantamento de contrastes já publicados, foram excluídos todos os contrastes que consideramos serem baseados em regionalismos ou num registro popular e/ou oral. Alguns dos que permaneceram foram complementados. Alguns novos contrastes, obtidos por vias empíricas, foram incluídos. Passamos então

para a segunda etapa, ou seja, para a comprovação e avaliação quantitativa dos contrastes encontrados, em *corpora*. Ao fazê-lo, não só pudemos aperfeiçoar e corrigir a descrição feita, como nos deparamos com novas diferenças que, por sua vez foram incluídas na lista anterior.

Para a comprovação e avaliação quantitativa dos contrastes analisamos separadamente dois *corpora* paralelos, compreendendo cerca de vinte mil palavras ao todo, criados de acordo com os nossos objetivos: o *corpus* de textos jornalísticos e o *corpus* de textos técnicos.

O *corpus* de textos jornalísticos foi criado a partir de textos coligidos em publicações brasileiras, cujo conteúdo temático incide sobre economia agrícola, administração/gestão de empresas e esporte/desporto, contendo cerca de 4.400 palavras. Estes textos foram convertidos para o Português Europeu por portugueses de maneira a formar um *corpus* paralelo.

Para o *corpus* de textos técnicos, partimos de originais em Inglês, recolhidos de manuais de instrução e artigos na área industrial, comercial e comunicações, totalizando cerca de 5.800 palavras. Estes textos foram traduzidos paralelamente por um tradutor brasileiro e por um tradutor português, formando assim um *corpus* paralelo para PE e PB.

Assim, o *corpus* de textos jornalísticos permite contagens de uma grande precisão, uma vez que na adaptação dos originais brasileiros para o PE, substituiu-se apenas o estritamente necessário para excluir todas as expressões de uso exclusivo da variante brasileira que continha, inclusive a nível de construção. Este método, embora ganhe muito em termos de precisão, perde um pouco em termos de naturalidade da linguagem.

Já o *corpus* de textos técnicos, também com cerca de cinco mil palavras para cada variante, não permite contagens tão rigorosas, uma vez que duas traduções do mesmo texto por duas pessoas diferentes, apresenta não só as variações dialetais, mas também variações devidas ao estilo pessoal de cada tradutor, interpretações diferentes, etc. Este *corpus* tem, entretanto, a vantagem de possuir maior autenticidade

Em consequência dessa diferença na construção dos *corpora*, a comparação dos resultados quantitativos entre os dois permitiu iluminar certos aspectos dificilmente detectáveis de outra forma. Por exemplo, enquanto no *corpus* de textos de imprensa o número de clíticos em cada uma das variantes era sensivelmente o mesmo, a análise do *corpus* de textos técnicos indica um uso muito mais significativo de clíticos no Português Europeu — em PB aparece apenas um terço dos que aparecem em PE.

Para proceder às contagens e comparações utilizamos ferramentas desenvolvidas no Grupo de Linguagem Natural do INESC (Medeiros, 1992), em particular para a anotação morfológica dos *corpora*, e para a localização de construções específicas em contexto.

Antes de passar para a demonstração de alguns exemplos, podemos adiantar que a primeira conclusão deste estudo foi a de que, genericamente há uma discrepância global de cerca de dez por cento entre as duas variantes, contando apenas contrastes morfológicos e lexicais. Este percentual foi calculado sobre o total de formas distintas não lematizadas do *corpus*, a partir do número de formas distintas contrastivas. Ou seja, as 4 400 palavras do *corpus* de textos de imprensa correspondem a pouco mais de 1 600 formas distintas não lematizadas. Aproximadamente 149 dessas formas distintas eram diferentes nas duas variantes, correspondendo 9,31% de contrastes. Para este cálculo foram considerados apenas

contrastes lexicais e morfológicos, não tendo sido possível, portanto, considerar também os contrastes sintáticos este efeito.

Não poderíamos acabar sem deixar de referir a nossa consciência de, pelo menos, duas fragilidades na base desta investigação. Uma é o conceito de língua padrão, cuja definição, por não ser pacífica, exigiria uma discussão que não caberia no âmbito deste trabalho. A outra é o problema da representatividade dos *corpora* que serviram de base à investigação, por sua limitada dimensão, tanto quantitativa quanto temática.

A título de exemplo, vamos examinar mais detalhadamente alguns dos resultados obtidos, nomeadamente os contrastes lexicais, e, na área morfo-sintática, a omissão de artigos.

1. Diferenças lexicais

Para o levantamento dos contrastes lexicais entre os dois textos de cada um dos *corpora*, consideramos como formas lexicais apenas as palavras pertencentes às seguintes classes gramaticais: substantivos, verbos, adjetivos e advérbios

1.1. Diferenças globais

1.1.2. Corpus de textos jornalísticos

O total de formas lexicais em cada variante do *corpus* de textos jornalísticos é de 2 270 formas, das quais 144 são formas contrastivas, correspondendo a 6,3%

Após lematizar o total de formas lexicais, ficamos com 1007 formas distintas lematizadas, das quais 70 são contrastivas, correspondendo a 7%

1.1.2. *Corpus* de textos técnicos

O total de formas lexicais em cada variante do *corpus* de textos técnicos é de 3.120 formas em PB e 3.107 em PE, somando 209 as formas contrastivas, o que corresponde a um percentual de 9,2%.

Uma vez lematizados cada um dos totais acima e eliminadas todas as repetições, ficaram 753 formas distintas lematizadas em PB e 942 em PE. Destas, 79 formas eram contrastivas, correspondendo a 9,8%.

Uma primeira ilação que se pode tirar daqui é o fato de no *corpus* de textos técnicos haver uma muito maior repetição de termos lexicais do que no *corpus* de textos jornalísticos.

1.2. Distribuição das diferenças lexicais

Aprofundando a análise de maneira a rentabilizar os resultados, ou seja, de fornecer dados mais detalhados segundo as exigências do desenvolvimento de um sistema de tradução automática simultaneamente para as duas variantes do Português, as palavras contrastivas foram divididas, por um lado segundo o tipo de diferenças e, por outro, segundo as categorias gramaticais a que pertencem.

1.2.1. Tipos de diferenças

Como *tipos de diferenças* consideramos aqui apenas (a) palavras com diferenças ortográficas, de que são distinguidas as (b) palavras com diferente acentuação, e as (c) palavras diferentes. Na classificação *palavras diferentes* foram englobados todas as palavras contrastivas, excetuando aquelas que se distinguem somente por sua ortografia (e acentuação). Assim, fazem parte deste item, tanto palavras que existem apenas em uma das variantes, quanto palavras que, existindo nas duas variantes, têm significados diferentes, ou em que pelo menos um dos seus significados é diferente (neste último caso, foi considerado

apenas o significado com que a palavra foi utilizada no contexto estudado). Inclui ainda, palavras que, existindo em ambas as variantes sofrem, no entanto, uma frequência de uso diferentes, como é o caso de *xícara* e *chávena*.

Ex.: - acentuação: PB: *comitê* PE: *comité*
 - ortografia: PB: *atividade* PE: *actividade*
 - palavras diferentes: PB: *goleiro* PE: *guarda-redes*

Corpus de textos jornalísticos:

Tipos de diferenças	no total das formas	nas formas distintas
acentuação	9 (6,2%)	7 (10%)
ortografia	68 (47,2%)	29 (41,4%)
palavras diferentes	67 (46,5%)	34 (48,5%)
Total	144	70

Corpus de textos técnicos:

Tipos de diferenças	no total das formas	nas formas distintas
acentuação	9 (4,3%)	7 (8,8%)
ortografia	50 (23,9%)	25 (31,6%)
palavras diferentes	150 (71,7%)	47 (59,5%)
Total	209	79

Como se pode observar, os contrastes relativos à acentuação são os que aparecem em número mais reduzido em ambos os *corpora*, mantendo-se também com percentuais bastante próximos.

Já o percentual de contrastes relativos à ortografia (excluída a acentuação) e o de contrastes relativo a *palavras diferentes* é praticamente igual no *corpus* de textos jornalísticos, mas distancia-se significativamente no *corpus* de textos técnicos. Visto que no primeiro o percentual de *palavras diferentes* é de 46,5%, enquanto no segundo é de 71,7%, podemos inferir que, em princípio, o número de contrastes do tipo *palavras diferentes* em terminologias específicas é significativamente superior aos contrastes lexicais do vocabulário geral. Observe-se ainda que em formas distintas, embora a diferença percentual de *palavras diferentes* entre os dois *corpora* tenha diminuído, mantém-se digna de nota.

1.2.2. Distribuição por categorias gramaticais

Dividindo os contrastes segundo suas categorias gramaticais, verifica-se que o maior número de contrastes encontra-se entre os substantivos nos dois *corpora*. É interessante notar que o segundo maior número de contrastes é de adjetivos no *corpus* de textos jornalísticos e de verbos no *corpus* de textos técnicos, embora o número global de adjetivos e verbos não difira muito entre os dois *corpora* (cf. o quadro abaixo).

Corpus de textos jornalísticos:

Categorias Gramaticais	Formas Totais	Formas Distintas
substantivos	114 (10,6% de 1.075)	50 (10,1% de 493)
verbos	9 (1,5% de 605)	7 (2,9% de 242)
adjetivos	17 (5,1% de 330)	11 (5,6% de 194)
advérbios	4 (1,5% de 265)	2 (2,5% de 8)

Corpus de textos técnicos:

categorias Gramaticais	no total das formas	nas formas distintas
substantivos	131 (7,7% de 1.690)	33 (7,8% de 421)
verbos	51 (5,8% de 875)	26 (10,2% de 253)
adjetivos	18 (5,1% de 350)	14 (7,3% de 190)
advérbios	9 (4,4% de 205)	6 (7,7% de 78)

2. Omissão de artigos

A omissão de artigos em diferentes estruturas no PB já foi abordada em vários estudos (ver Roberts e Kato, 1993). Entretanto, parece haver uma enorme oscilação entre omissão e realização do artigo no PB, que depende não só das regiões ou dialetos, como também do estilo pessoal de cada autor ou idioletos. Nos *corpora* comparados encontramos omissão de artigos no PB antes de pronomes possessivos, depois de pronomes indefinidos e em expressões de generalização. Não foram encontrados casos de omissão de artigo antes de nome próprio, embora seja sabido que o fenômeno é freqüente em PB. A única possibilidade que pudemos captar de omissão de artigos em PE dá-se antes de alguns topônimos, mas não ocorre nenhuma vez nos *corpora* examinados.

2.1. Antes de pronomes possessivos

O *corpus de textos jornalísticos* contém um total de 34 pronomes possessivos em cada uma das variantes. No PB 26 destes não são precedidos por artigo, ou seja, 76,47%. No PE apenas 4 não são precedidos por artigo, ou seja 11,76%. Reduzidos a formas distintas, os pronomes possessivos não precedidos por artigo somam 5. Estes números incluem obviamente os artigos em contração com preposições

ex.: PB: "Este, antes de tudo, tem prazer com seu trabalho..."

PE: "Este, antes de tudo, tem prazer com o seu trabalho..."

O *corpus* de textos técnicos contém um total de 17 pronomes possessivos em PB, dos quais 2 não são precedidos por artigo, ou seja 11,76%. Em PE há um total de 18 pronomes possessivos, não havendo nenhuma omissão de artigo, ou seja, 0%.

ex.: PB: "... você vende seus produtos..."

PE: "... contacte o seu representante..."

A discrepância dos resultados entre os dois *corpora* (76,47% e 11,76%) permite-nos avaliar a dimensão atual de flutuação do fenómeno em PB.

2.2. Depois de pronomes indefinidos

Foi encontrado apenas um caso de omissão de artigo depois de pronome indefinido no *corpus* de textos jornalísticos:

PB: *todo mundo*

PE: *toda a gente*

Se bem que, neste caso, tratem-se na verdade de expressões diferentes, parece-nos que o fenómeno é relevante. Parece-nos que a omissão verifica-se, sobretudo, depois dos pronomes *todo* e *tudo*, como por exemplo em *tudo que ele disse* em PB e *tudo o que ele disse* em PE, *toda vez* em PB e *todas as vezes* em PE ou, ainda, *todo dia* em PB e *todo o dia/todos os dias* em PE.

2.3. Outros

No *corpus* de textos jornalísticos verificou-se um caso contrastivo em complemento nominal, que consiste na generalização semântica do nome em PB e a sua particularização em PE através do artigo:

PB: "... tarefas planejadas pelo administrador de sistema."

PE: "... tarefas planeadas pelo administrador do sistema."

Este caso único seria de pouca consistência se tentássemos a partir daí deduzir ilações determinantes. Considerando, porém, outros exemplos recolhidos noutros contextos como, por exemplo, *Trabalhadores de construção civil bloquearam a estrada* em PB e *Os trabalhadores da construção civil bloquearam a estrada* em PE, parece-nos que este caso ilumina um campo de investigação ainda não explorado nas publicações sobre a especificidade do PB em relação ao PE — ao menos nos estudos a que tivemos acesso.

2.4. Tipos de omissão não representados neste corpus

A omissão de artigos realiza-se, ainda, em dois outros casos que não estão representados nos *corpora* examinados. Trata-se da possibilidade de omissão do artigo antes de alguns topônimos em PE e da possibilidade de omissão do artigo antes de nomes próprios e nomes com função de sujeito.

2.4.1. Antes de topônimos

A freqüente omissão do artigo antes de alguns topônimos em PE, nomeadamente, *em Inglaterra*, *em Itália*, *em França*, *em Espanha*, *em África*, enquanto em PB seria *na Inglaterra*, *na Itália*, *na França*, *na Espanha* e *na África*, tem sido referida em algumas gramáticas (ver, p.ex., Abreu e Benamor, 1994, e Cunha e Cintra, 1987). Dieter Woll (1993) num cuidadoso estudo sobre o assunto verifica que a omissão também se verifica com os topônimos Mônaco e Luxemburgo, ainda que com menor freqüência.

2.4.2. Antes de nome próprio

É freqüente, ainda, a omissão do artigo antes de nome próprio em PB. A frase *Pedro saiu*, por exemplo, é possível em PB, enquanto em PE seria obrigatoriamente *(O) Pedro saiu*.

Referências Bibliográficas

- ABREU, Helena, BENAMOR, Rita, adapt. dos exercícios para Português Brasileiro por WITTMANN, Luzia Helena. *Gramática del Portuguese Moderno*, Zanichelli Editore, Bologna, 1994.
- BIBER, Douglas. "Textual Comparison of British and American Writing" in *American Speech* 2:99-119, 1987.
- BIDERMAN, Maria Tereza C. *Vocabulário Fundamental: Cultura e Sociedade*, UNESP, Araraquara, SP, exemplar policopiado.
- CUESTA, Pilar Vázquez, LUZ, Maria Albertina Mendes da. *Gramática da Língua Portuguesa*. Edções 70, Lisboa, 1971.
- CUNHA, Celso e CINTRA, Lindley. *Nova Gramática do Português Contemporâneo*, Edições João Sá da Costa, Lisboa, 1987.
- CYRINO, Sonia Maria L. "Observações sobre a mudança diacrônica do sujeito nulo no Brasil: objeto nulo e clíticos", in Ian Roberts and Mary A. Kato, *Português Brasileiro, uma viagem diacrônica*, Ed. UNICAMP, Campinas, SP, 1993, pp. 163-184.
- DUARTE, Maria Eugênia L. "Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil", in Ian Roberts and Mary A. Kato, *Português Brasileiro, uma viagem diacrônica*, Ed. UNICAMP, Campinas, SP, 1993, pp. 107-128.
- GALVES, Charlotte C. "O enfraquecimento da concordância no português brasileiro", in Ian Roberts and Mary A. Kato, *Português Brasileiro, uma viagem diacrônica*, Ed. UNICAMP, Campinas, SP, 1993, pp. 387-408.
- HOLANDA, Aurélio Buarque de. *Dicionário Aurélio Eletrônico*, Editora Nova Fronteira, Rio de Janeiro, 1993.
- KATO, Mary A. Kato. "Recontando a história das relativas em uma perspectiva paramétrica", in Ian Roberts and Mary A. Kato (org.), *Português Brasileiro, uma perspectiva diacrônica*, Ed. UNICAMP, Campinas, SP, 1993, pp. 223-262.
- LUFT, Celso Pedro. *Dicionário Prático de Regência Verbal*, Editora Ática, São Paulo, 1987.
- MARTIN, Harris and NIBEL, Vincent. *The Romance Language*, Routledge, London, 1988.
- MATEUS, Maria Helena Mira, et alii. *Gramática da Língua Portuguesa*, 3ª ed. refundida, 1971.

- MEDEIROS, José Carlos. "Ferramentas de Manipulação de Corpora". in SANTOS, Diana (ed.), *Processamento de Corpora de Texto no INESC*, relatório INESC RT/65-92, Lisboa, 1992.
- MEDEIROS, José Carlos, MARQUES, Rui, SANTOS, Diana. "Português Quantitativo", in *Actas do 1º Encontro de Processamento da Língua Portuguesa Escrita e Falada, EPLP'93*, Lisboa, 1993, pp. 33-38.
- MORAIS, Clóvis Barleta de. *Diferenças entre a Sintaxe do Português Clássico e a do Moderno*, UNESP, Araraquara, SP, s/d, exemplar policopiado.
- PRATA, Mário. *Dicionário de Português - Schifaitzfoivre*, Editora Globo S.A., São Paulo (SP), 1993.
- ROBERTS, Ian. "O português brasileiro no contexto das línguas românicas". in Ian Roberts and, Mary A. Kato (org.), *Português Brasileiro, Uma Viagem Diacrônica*, Ed. UNICAMP, Campinas, SP, 1993, pp. 409-425.
- TARALLO, Fernando. "Diagnosticando uma Gramática Brasileira: o português d'aquém e d'além-mar ao final do século XX", in Ian Roberts and Mary A. Kato (orgs.), *Português Brasileiro, Uma Viagem Diacrônica*, Editora UNICAMP, Campinas, SP, 1993, pp. 69-106.
- TEYSSIER, Paul, trad. de CARVALHO, Margarida Chorão. *Manual de Língua Portuguesa (Portugal-Brasil)*, Coimbra Editora, 1989.
- VILAR, Mauro. *Dicionário Contrastivo Luso-Brasileiro*, Editora Guanabara, Rio de Janeiro, 1989.
- WITTMANN, Luzia Helena, PEREIRA, Maria de Jesus. *European and Brazilian Portuguese: Survey of Morphological, Syntactical and Orthographic Differences*, Relatório INESC, Lisboa, 1994.
- WITTMANN, Luzia Helena, PEREIRA, Maria de Jesus. *European and Brazilian Portuguese: Quantitative Report of Lexical, Syntactical and Orthographic Differences*.
- WOLL, Dieter, "Na França e em França, mas só: na Alemanha - porquê? Pequena contribuição para a história do artigo definido nas línguas românicas", in Jürgen Schmidt-Radefeldt (ed.), *Semiótica e linguística portuguesa e românica*, Gunter Narr Verlag Tübingen, 1993.