

# **Integrando a variação social e métodos quantitativos na investigação sobre linguagem e cognição: para uma sociolinguística cognitiva do português europeu e brasileiro**

Integrating social variation and quantitative methods in the investigation of language and cognition: towards a cognitive sociolinguistics of European and Brazilian Portuguese

Augusto Soares da Silva  
Universidade Católica Portuguesa – Braga

## **Abstract**

The purpose of this paper is two-fold. First it argues the case for systematic integration of language variation and corpus-based methodology with its accompanying methods of quantitative and statistical analysis in the cognitive study of language, as it has been occurring in Cognitive Linguistics. The second and most important goal is to present a program for Cognitive Sociolinguistics of European and Brazilian Portuguese, centered in the issue of convergence and divergence between the two national varieties in the past 50 years. First, we will highlight two instances of tension in the current stage of the development of Cognitive Linguistics (one focus of tension on the object and the other on the methodology). Then we will show how its usage-based perspective implies to integrate and to correlate a social object and an empirical methodology. We will then identify specific contributions of the emergent Cognitive Sociolinguistics and onomasiological profile-based methodology to the study of language variation. Finally, we will show the results of our sociolexicological research about convergence and divergence in the vocabulary of

football and clothing – the two national varieties converge for football and diverge for clothing – and we point out the extension of this first stage of research to the domains of functional words and non-lexical variables.

### Keywords

Cognitive Linguistics; Cognitive sociolinguistics; Corpus linguistics; Denotational synonyms; European/Brazilian Portuguese; Language variation; Methodology; Quantitative methods; Social cognition; Socilectometry; Usage-based model; Statistics.

### Resumo

Este artigo tem um objectivo duplo. Um é argumentar sobre a necessidade de integrar sistematicamente a variação linguística e a metodologia do *corpus* acompanhada de métodos quantitativos e de análise estatística no estudo cognitivo da linguagem, tal como tem sido desenvolvido pela Linguística Cognitiva. O outro, e mais importante, é apresentar um programa de sociolinguística cognitiva do português europeu e brasileiro centrado na questão da convergência e divergência entre as duas variedades nacionais nos últimos 50 anos. Primeiramente, assinalaremos duas tensões no desenvolvimento actual da Linguística Cognitiva (uma no objecto e outra no método) e mostraremos como é que a sua perspectiva *baseada no uso* implica integrar e complementar o objecto social e a metodologia empírica. A seguir, identificaremos os contributos específicos da emergente Sociolinguística Cognitiva e do método dos perfis onomasiológicos para o estudo da variação linguística. Finalmente, apresentaremos os resultados da nossa investigação sociolexicológica sobre convergência e divergência nos vocabulários do futebol e do vestuário – as duas variedades convergem no futebol e divergem no vestuário – e indicaremos as extensões desta primeira fase da investigação para os domínios das palavras funcionais e das variáveis não-lexicais.

### Palavras-chave

Cognição social; Estatística; Linguística Cognitiva; Linguística de *corpus*; Metodologia; Métodos quantitativos; Modelo baseado no uso; português europeu/brasileiro; Sinónimos denotacionais; Socilectometria; Sociolinguística cognitiva; Variação linguística.

## 1. Introdução

Para além do epíteto que lhe dá o nome e a define como o movimento mais consistente e consequente de estudo *cognitivo* da linguagem, a Linguística Cognitiva apresenta-se também como um *modelo baseado no uso* (LANGACKER, 1988, 2000) e um modelo *(re)contextualizante* da linguagem, no sentido de (re)introduzir no estudo das línguas as várias facetas do *contexto*, que foram eliminadas pelo programa autonomista da Gramática Gerativa. Tais características implicam, já por si, uma atenção particular pela variação linguística e um interesse preferencial por métodos empíricos. Todavia, e por um lado, a filosofia *experientialista* da Linguística Cognitiva (LAKOFF & JOHNSON, 1999), em particular a tese da *corporização* do pensamento e da linguagem, tende a conceber a cognição humana como fenómeno individual e invariável ou universal; por outro lado, a mesma filosofia não-objectivista e a assunção da centralidade do significado no estudo cognitivo das línguas mostram alguma relutância pelos métodos empíricos, por natureza, objectivistas das descrições linguísticas, tendendo a preferir a introspecção, como método de acesso directo aos fenómenos semânticos, por natureza menos objectivos do que quaisquer outros fenómenos linguísticos. Criam-se, assim, duas tensões no desenvolvimento actual da Linguística Cognitiva: uma *tensão no objecto* da pesquisa entre o apelo a todas as dimensões contextuais da linguagem, como dimensões da experiência humana, incluindo os aspectos sociais, culturais e variacionais, e a tradicional concepção individual e neurológica da cognição humana; e uma *tensão no método* da pesquisa entre o interesse cada vez maior pelos métodos empíricos, sejam experimentais, neurofisiológicos ou de *corpus* (ver GONZALEZ-MARQUEZ *et al.*, 2007), e um certo cepticismo teórico pela metodologia objectivista no estudo de fenómenos subjectivamente experienciados, como são os linguísticos em geral e os semânticos em particular.

Procuramos neste estudo argumentar, continuando a posição construtiva de alguns autores, como Geeraerts (2005, 2006), sobre a importância ou, melhor,

a inevitabilidade de integrar sistematicamente a variação social e métodos quantitativos baseados em *corpora* no estudo cognitivo das línguas. O segundo e mais importante objectivo deste artigo é apresentar, no quadro da emergente Sociolinguística Cognitiva (KRISTIANSEN & DIRVEN, no prelo), um programa de sociolinguística cognitiva do português europeu e brasileiro centrado na questão da convergência ou divergência entre as duas variedades nacionais. Nas duas primeiras secções, dedicadas às duas tensões referidas, exporemos as razões para integrar o objecto social e a metodologia empírica na perspectiva da Linguística Cognitiva, já teoricamente favorável a estes desenvolvimentos. Nas duas secções seguintes, identificaremos os contributos específicos da sociolinguística cognitiva para o estudo da variação linguística e apresentaremos o método da variação onomasiológica envolvendo sinónimos denotacionais e os métodos quantitativos para medir a convergência e a divergência entre as duas variedades nacionais do português. Nas duas últimas secções, sumariaremos os resultados da análise sociolexicológica sobre o vocabulário do futebol e o vocabulário do vestuário e indicaremos as extensões desta primeira fase da investigação para os domínios das palavras funcionais, particularmente preposições, e das variáveis gramaticais.

## 2. Cognitivo e social

Muito sinteticamente, podemos definir a Linguística Cognitiva como um modelo recontextualizador, experiencialista, orientado para o significado e baseado no uso. Cada uma destas quatro características implica uma investigação socialmente orientada. O mesmo é dizer que, relativamente à tensão no objecto da pesquisa, anteriormente mencionada, temos aqui quatro razões intrínsecas para integrar sistematicamente o *social* no *cognitivo* e, num plano mais geral, entender a cognição como *cognição social*.

Como modelo *recontextualizador*, em resposta à descontextualização da gramática, exemplarmente efectuada pela Gramática Generativa, a Linguística Cognitiva não poderá deixar de incorporar nenhuma das facetas do *contexto*: não só as bases cognitivas e experienciais dos falantes ou contexto cognitivo, como também o nível interaccional do uso linguístico ou contexto situacional e o ambiente sociocultural da língua ou contexto social. A integração sistemática dos aspectos sociais da linguagem nas diversas linhas de investigação cognitiva das línguas constituirá a fase final da expansão da Linguística Cognitiva.

Como modelo *experencialista*, a Linguística Cognitiva deverá atender a todas as dimensões da experiência humana, construídas pela e na linguagem, não só a experiência individual (corpórea, neurofisiológica), como também a experiência colectiva, social e cultural e com esta naturalmente as diferenças entre culturas, grupos sociais ou mesmo indivíduos. Como veremos mais adiante, esta implicação coloca o desafio de reinterpretar a cognição como *cognição social* (ou *cognição situada*) e especificamente saber como é que interagem factores individuais, neurofisiológicos e universais, de um lado, e factores interindividuais, socioculturais e variacionais, do outro.

Como modelo *orientado para o significado*, a Linguística Cognitiva não poderá menosprezar a variação sociolinguística, por duas ordens de razões mais específicas: por um lado, os traços semânticos não-ortodoxos da *saliência* e *vaguidade*, amplamente explorados em Semântica Cognitiva, contribuem para a variação linguística; por outro lado, a variação sociolinguística constitui uma forma específica de significado. Efectivamente a variação linguística dá origem a diferentes tipos de significado não-denotacional: o significado *emotivo* (de termos pejorativos, por exemplo), o significado *estilístico* e mais estritamente *sociolinguístico* (de termos populares ou eruditos, regionais ou sociais) e o significado *discursivo* (único de expressões como as interjeições e os marcadores discursivos; presente em termos como *senhor*, *você*, *tu* e outras formas de tratamento).

Finalmente, como *modelo baseado no uso* – caracterização de Langacker (1988, 2000), popularizada por Barlow & Kemmer (2000) e reforçada em publicações mais recentes como as de Tomasello (2000, 2003), Bybee & Hopper (2001), Croft & Cruse (2004) ou Geeraerts (2006) –, a Linguística Cognitiva não pode evitar a variação linguística como seu objecto de investigação. A razão é simples: a variação é a consequência imediata e inevitável do uso; nunca uma comunidade linguística é totalmente homogénea e qualquer língua é um *diassistema* social. Por outro lado, o interesse crescente pelos métodos empíricos, implicado por esta mesma razão da importância do uso, como veremos na secção seguinte, obriga o linguista a ter em conta os aspectos variacionais. Na verdade, um *corpus* representativo nunca será absolutamente homogéneo, incluindo sempre alguma variação *lectal*, isto é, relacionada com qualquer tipo de variedades ou variantes de uma língua ou, simplesmente, *lectos* (dialectos, regiolectos, sociolectos, idiolectos, variedades nacionais, registos ou estilos, etc.). Mesmo que a análise não se centre na variação lectal, esta será sempre um factor,

no sentido de que é necessário saber se a variação observável no *corpus* resulta ou não de factores lectais. Não há, portanto, maneira de evitar a variação linguística a partir do momento em que se assumia séria e plenamente um modelo baseado no uso. Como alerta Geeraerts (2001, p. 53), “it is impossible to take seriously the claim that Cognitive Linguistics is a usage-based approach and at the same time to neglect the social aspects of language use”.

Se a Linguística Cognitiva está teoricamente bem colocada para integrar no objecto da pesquisa linguística os aspectos sociais da linguagem e sendo certo que a investigação socialmente orientada e a investigação variacional não estão ausentes do seu programa, como veremos na secção 4, não é menos verdade que o estudo sistemático da variação intralinguística é ainda uma tendência minoritária da sua agenda. Além disso, ou como causa disso, ela está confrontada com a existência no seu seio de dois modos, em conflito, de definir experiencialismo e cognição, e ainda não resolveu a questão de saber como é que interagem factores individuais e interindividuais.

As estruturas linguísticas exprimem conceptualizações e as conceptualizações realizadas na e através da linguagem estão intrinsecamente relacionadas como o modo como os seres humanos experienciam a realidade, tanto fisiológica como culturalmente. Constitui este *experiencialismo* a própria filosofia da Linguística Cognitiva, paradigmaticamente elaborada por Lakoff & Johnson (1999), em ruptura com o *objectivismo* de outras grandes correntes linguísticas, com a estruturalista e a generativista. Mas a interpretação de Lakoff & Johnson (1999), com a qual mais se identificou a Linguística Cognitiva no seu conjunto, é parcial: a tese da *corporização* (“embodiment”) do pensamento e da linguagem ou a *filosofia na carne* foca a vertente individual e universal da cognição humana (o corpo é um universal da experiência humana), o seu lado físico e neurofisiológico, recentemente explorado por Lakoff (2003) na sua Teoria Neural da Metáfora. Ora, tendo a experiência humana uma dimensão também colectiva e interactiva, impõe-se não reduzir a filosofia experiencialista e o princípio da corporização a operações neurais meramente individuais e reconhecer a natureza socialmente interactiva da linguagem e o seu ambiente cultural como elementos igualmente fundacionais da perspectiva cognitiva. É justamente isto que tem sido reclamado por autores como Tomasello (1999), Zlatev (1999, 2003), Sinha & Jensen de López (2000), Harder (2003), Bernárdez (2004, 2005). Só assim é que a Linguística Cognitiva poderá evitar cair no perigo do solipsismo epistemológico ou no que Sinha (1999) designa como “solipsismo neural”.

O problema está então na própria concepção da cognição humana e na separação entre o individual e o colectivo, o interno e o externo, o pensamento e a acção, a cognição e a actividade – pensamento individual e acção colectiva, será esta a equação adequada? Poderemos conceber a cognição sem a interacção? Poderemos continuar a assumir, como é habitual na cultura ocidental (e não só), que o pensamento individual interno tem algum tipo de preeminência sobre a actividade supra-individual externa ou sobre o pensamento dirigido para a (inter)acção? Vários filósofos, psicólogos, neurocientistas e linguistas respondem que não. Lembremos a inseparabilidade da cognição e da emoção, demonstrada nos estudos de Damásio (1995, 2000). Efectivamente, tem havido, nos últimos vinte ou mais anos, um alargamento significativo do âmbito da cognição: desde uma perspectiva puramente interna, com a primeira geração das ciências cognitivas, à perspectiva *corporizada* (VARELA, THOMPSON & ROSCH, 1991; EDELMAN, 1992; DAMÁSIO, 1995, 2000; LAKOFF & JOHNSON, 1999; GIBBS, 2006) aberta ao exterior e, mais recentemente, à inclusão da situação, actividade ou interacção na cognição e, assim, à noção de *cognição situada* ou *cognição social* (ZLATEV, 1997, 1999; TOMASELLO, 1999; BERNÁRDEZ, 2004, 2005; e, particularmente, as obras colectivas organizadas por ZIEMKE, ZLATEV & FRANK e FRANK, DIRVEN, ZIEMKE & BERNÁRDEZ, ambas ainda no prelo).

Bernárdez (2005) sumaria importantes pontos da concepção sócio-cognitiva da linguagem: as línguas existem somente na forma de actividade social; a actividade linguística é essencialmente colectiva, social por natureza; conseqüentemente, a linguagem é um fenómeno inerentemente histórico e, como tal, fenómenos que hoje são examinados a nível individual (como a metáfora e a metonímia) são o resultado de uma cristalização social, histórica; a actividade linguística (o uso linguístico) determina formas e estruturas linguísticas, umas preferidas e outras não; através de um processo de integração cognitiva, as formas produzidas e preferidas pela actividade linguística são gradualmente fixadas na mente do indivíduo, a ponto de a sua conexão original com a actividade se perder e, desse modo, se tornarem parte da cognição individual – é assim que a actividade linguística social influencia os próprios processos cognitivos.

### 3. Cognitivo e empírico

Várias das características que, como vimos na secção anterior, definem a Linguística Cognitiva e a comprometem a incorporar sistematicamente no objecto da pesquisa linguística a variação social, também a conduzem a optar pela metodologia empírica.

O seu epíteto de (ciência) *cognitiva* leva a Linguística Cognitiva a seguir os métodos empíricos que se têm revelado adequados e produtivos no paradigma das ciências cognitivas em geral, designadamente as técnicas experimentais da psicologia, as modelações computacionais e a análise neurofisiológica. Gibbs (1994, 2006) e Lakoff (2003) são, entre outros, excelentes exemplos da aplicação de métodos experimentais e neurofisiológicos, respectivamente, no estudo da metáfora conceptual. Há, porém, que notar que esta ideia de que o cognitivo implica o empírico pode não se verificar. A prova é que a Gramática Generativa, que tomou parte na chamada revolução cognitiva dos anos 50, sendo portanto também uma “linguística cognitiva”, adoptou a metodologia oposta – essencialmente introspectiva. Mas há que reconhecer que o adjectivo *cognitivo* não significa a mesma coisa nos dois quadros teóricos quer linguísticos quer das ciências cognitivas em geral: na Gramática Generativa esse epíteto significou a tese da autonomia da linguagem relativamente a outras capacidades cognitivas (e no cognitivismo clássico significou o objectivismo), ao passo que na Linguística Cognitiva isso significa justamente o contrário, isto é, que a linguagem é parte integrante, e não um módulo separado, da cognição e que, por isso mesmo, deve ser estudada no contexto de outras capacidades cognitivas. Ora, é esta concepção não-autonomista da capacidade cognitiva da linguagem que propicia a adopção da metodologia empírica.

Como segunda característica da Linguística Cognitiva que favorece a adopção de uma metodologia empírica, está a sua natureza de *modelo baseado no uso* – uma orientação de baixo-para-cima, maximalista e não-redutiva, em contraste com a perspectiva de cima-para-baixo, minimalista e redutiva da Gramática Generativa. A razão é simples: não se pode ter uma linguística baseada no uso se não se estudar o uso efectivo da língua, e o uso efectivo da língua manifesta-se em dados espontâneos e não-eliciados de um *corpus* ou em dados eliciados de inquéritos, tarefas de resolução de problemas ou outras experimentações.

Uma terceira característica da Linguística Cognitiva que a conduz à investigação empírica é a sua concepção *contextualizada* da linguagem. O interesse pelas diferenças sociais e culturais leva-a ao interesse pela variação linguística e o interesse pela variação sociolinguística leva-a inevitavelmente a uma metodologia baseada em dados do *corpus*. Eis como a implicação, vista na secção anterior, do cognitivo para o social traz consigo esta implicação do cognitivo para o empírico: o reconhecimento dos aspectos sociais da linguagem implica uma metodologia empírica de observação do uso dos dados e a forma mais natural de os encontrar é em *corpora* textuais representativos.

Há, todavia, uma outra característica essencial da Linguística Cognitiva que pode favorecer a via metodológica oposta: a ênfase posta no significado linguístico joga com a ideia de que a introspecção é o único método de acesso directo aos fenómenos semânticos. Como os significados não se dão directamente nos dados de um *corpus* nem nos dados eliciados, ganha (mais) força a ideia de que a introspecção é, em última instância, o método apropriado para os estudar. Continuando a ideia, uma metodologia objectivista não serve para uma teoria não-objectivista. Ora, temos aqui um equívoco, que o exemplo da Psicologia ensinaria a evitar, ao demonstrar que se pode e deve estudar objectivamente fenómenos não objectivos como são os cognitivos. Ou seja: os processos cognitivos que se estudam em nível do objecto podem ser não-objectivos, mas a descrição no meta-nível teórico tem que ser objectiva.

Outros factores também podem explicar esta resistência da Linguística Cognitiva à metodologia empírica. Um deles é o próprio exemplo de algumas das principais figuras e fundadores da Linguística Cognitiva: Langacker (1987, 1991), Talmy (2000) ou Fauconnier (1997, FAUCCONNIER & TURNER, 2002) aderem, se não na teoria pelo menos na prática, ao método tradicional da introspecção (ver TALMY, 2007, para uma defesa da introspecção); não há nos seus estudos inovadores e inspiradores nenhuma análise sistemática de *corpus* nem nenhuma experimentação desenvolvida. Já Lakoff, mas mais recentemente, tem procurado evidência neurofisiológica para a sua teoria da metáfora conceptual (LAKOFF, 2003). Observe-se, a propósito, uma certa divergência dentro da Linguística Cognitiva entre o ramo europeu, mais adepto da metodologia empírica de *corpus*, e o ramo americano, mais inclinado para a metodologia introspectiva, qual herança da Gramática Generativa. Um outro factor está no tradicional cepticismo teórico em relação aos métodos empíricos, sobretudo quando estão em causa fenómenos

que não são acessíveis a partir “de fora”, que não podem ser tomados objectivamente, como são os fenómenos semânticos e todos os outros fenómenos cognitivos. E não só aquele cepticismo mais forte que considera que, face à superioridade da interpretação, a evidência empírica não é necessária; como também o cepticismo mais moderado que considera que a evidência empírica, por mais relevante que seja, não é suficiente, pois há fenómenos subjectivamente experienciados que se lhe escapam.

Num estimulante e construtivo artigo sobre metodologia em Linguística Cognitiva, Geeraerts (2006) assevera que a linguística actual, em geral, e a Linguística Cognitiva, em particular, precisam de uma *revolução empírica*. Os estudos linguísticos do século passado foram pródigos em modelos e grandes teorias. É tempo, no início deste novo século, de encontrar provas para refinar ou abandonar essas teorias. O que falta é essencialmente o trabalho de operacionalização empírica das hipóteses linguísticas. Para esse efeito, a linguística tem ao seu dispor vários métodos empíricos: pode utilizar as técnicas de experimentação da psicologia, as técnicas da análise neurofisiológica, as técnicas topográficas e de trabalho de campo da sociologia e dispõe ainda da análise de *corpus*, que é o método empírico mais genuinamente linguístico.

Mas a necessária revolução empírica tem duas exigências específicas: sofisticação de técnicas e interpretação dos dados. Por um lado, não basta utilizar o *corpus* de um modo que Tummers, Heylen & Geeraerts (2005) classificam como investigação meramente *ilustrada por um corpus*, isto é, não basta dispor de um *corpus* como mero repositório de ocorrências e depois analisar essas ocorrências de modo predominantemente intuitivo. É de facto este o procedimento da maior parte dos estudos que se dizem baseados num *corpus*. É preciso fazer análise estatística; é preciso desenvolver e utilizar métodos de análise quantitativa e multivariacional. Exemplos de trabalhos com apreciável sofisticação quantitativa e estatística dentro da Linguística Cognitiva são os de Geeraerts, Grondelaers & Speelman (1999), Grondelaers, Speelman & Geeraerts (2002), Speelman, Grondelaers & Geeraerts (2003), Grondelaers, Geeraerts & Speelman (2007), Gries (2003), Stefanowitsch (2003), Stefanowitsch & Gries (2003, 2005). Mas trabalhos como estes são ainda uma minoria – ver o estado da arte em Tummers, Heylen & Geeraerts (2005), Gries & Stefanowitsch (2006) e Gonzalez-Marquez *et al.* (2007). Por outro lado, a análise de *corpus* nem é automática nem é arredada da tarefa hermenêutica. Quer isto dizer que é preciso

fazer o trabalho de operacionalização das hipóteses, isto é, formular hipóteses que sejam empiricamente testáveis, confrontar as hipóteses com os dados empíricos, refinar gradualmente as hipóteses confrontando-as repetidamente com os dados empíricos. Assim, e contrariando o equívoco várias vezes cometido, a investigação empírica não implica abandonar a teoria em favor de uma investigação puramente descritiva, mas antes testá-la e refiná-la.

Em suma, a *investigação empírica* envolve, como Gearererts (2006) claramente explica, as seguintes tarefas: recolha e colecção representativa de dados, métodos quantitativos e análise estatística, formulação de hipóteses, operacionalização de hipóteses e um *ciclo empírico* de tipo hélice de refinamento gradual das interpretações através do confronto repetido com os dados empíricos.

#### 4. Sociolinguística Cognitiva

O social e a variação linguística estão presentes em algumas linhas de investigação em Linguística Cognitiva, desde o seu início. É o caso do estudo dos *modelos culturais*, “teorias populares” ou *modelos cognitivos culturais*, isto é, representações individualmente idealizadas e interindividualmente partilhadas sobre fenómenos sociais e culturais como o casamento, as emoções, a mente humana, a linguagem humana, o género e as raças, os astros, concepções sociais e políticas, normas sociais e jurídicas, desde os trabalhos reunidos por Holland & Quinn (1987) até estudos mais recentes como os coligidos por Dirven, Frank & Pütz (2003). A investigação sobre estes e outros aspectos da variação cultural, incluindo as explorações das especificidades culturais da metáfora conceptual (LAKOFF, 1993; KÖVECSÉS, 2000, 2005) ou da cognição espacial (LEVINSON, 2003), constitui o que no seio do mesmo movimento se designa como *linguística cultural* (PALMER, 1996). A tudo isto devem ainda juntar-se os estudos sobre linguagem e ideologia, como os de Lakoff sobre política (sobretudo LAKOFF, 1996) e os mais recentes reunidos em Dirven, Hawkins & Sandikcioglu (2001) e Dirven, Frank & Ilie (2001). Estes estudos exploram as conotações negativas das ideologias e relações de poder, mas há também trabalhos sobre conotações positivas ou neutras, como o de Geeraerts (2003) sobre modelos culturais (racionalista e romântico) da variação e standardização linguísticas ou o de Kristiansen (2003) sobre estereótipos associados a variantes fonéticas como pontos de referência na identificação regional e social dos falantes.

A variação linguística tem sido estudada na perspectiva cognitiva (ou cognitivo-funcionalista) de três pontos de vista: diacrónico, com a extensa investigação sobre gramaticalização, mudança semântica e mudança fonológica (por ex., HOPPER & TRAUGOTT, 1993; TRAUGOTT & DASHER, 2002; GEERAERTS, 1997; BLANK & KOCH, 1999; BYBEE, 2001); tipológico e antropológico, incluindo os estudos referidos no parágrafo anterior e outros trabalhos como os de M. Bowerman, S. Levinson e D. Slobin sobre diferenças interlinguísticas; e psicolinguístico, sobre a aquisição e o desenvolvimento da linguagem, destacando-se o estudo de Tomasello (2003).

Mas a variação intralinguística ou diversidade sociolinguística, essa tem sido menos estudada dentro da Linguística Cognitiva. Representam esta linha de investigação os estudos exemplares de Geeraerts, Grondelaers & Speelman (1999) e Speelman, Grondelaers & Geeraerts (2003) sobre a variação lexical e com uma rica metodologia quantitativa e variacional de *corpus*, o trabalho de Kristiansen (2003) sobre variação fonética e a colectânea de estudos organizada por Kristiansen & Dirven (no prelo), institucionalizando a área emergente da *sociolinguística cognitiva*. Mais centrada na variação linguística, a Sociolinguística Cognitiva procura, no entanto, integrar as linhas anteriores de investigação socialmente orientada, examinando as correlações entre a variação linguística e modelos culturais, a variação linguística e a diversidade social e cultural, a variação linguística e ideologias e incluindo as questões de política da língua.

Cabe perguntar o que é que há de específico na emergente sociolinguística cognitiva e que contributo pode ela oferecer à investigação sociolinguística em geral? Podemos apontar três especificidades e contributos. Em primeiro lugar, a própria perspectiva cognitiva dos fenómenos variacionais; concretamente, a aplicação dos diversos modelos descritivos já existentes no estudo da variação linguística. Por exemplo, a abordagem da variação fonética pela teoria do protótipo (KRISTIANSEN, 2003) ou a utilização de metáforas conceptuais e modelos culturais na identificação de atitudes linguísticas (GEERAERTS, 2003). Em segundo lugar, a exploração da *cognição social*, particularmente, a elucidação da interacção dialéctica entre o nível individual cognitivo e o lado social das normas colectivas. Finalmente, mas não menos importante, o desenvolvimento de métodos quantitativos baseados em *corpora* e de métodos de análise multivariacional da confluência de factores conceptuais, discursivos e variacionais dos fenómenos linguísticos.

Entre os vários tipos de variação linguística, a variação lexical é a que tem sido objecto de maior estudo dentro da Linguística Cognitiva, pelo facto de que este paradigma começou a afirmar-se na área da semântica lexical. Os estudos de Semântica Cognitiva muito têm contribuído para o desenvolvimento da teoria lexicológica, em geral, e da teoria da variação lexical, em particular. Podem apontar-se dois importantes contributos. Por um lado, passa-se de um interesse nas estruturas lexicais (campos lexicais e taxionomias), típico da semântica estrutural, para um interesse no uso e em fenómenos de saliência ou prototipicidade. Por outras palavras, a teoria lexicológica alarga o seu objecto de pesquisa da dimensão *qualitativa* dos fenómenos lexicais (unidades e relações) para a dimensão *quantitativa* das diferenças de saliência conceptual entre essas unidades e relações. Por outro lado, faz-se a extrapolação dos efeitos de saliência do domínio semasiológico para o domínio onomasiológico, com a noção de *ancoragem* conceptual (“entrenchment”) entre diferentes categorias e a hipótese do *nível básico* das taxionomias lexicais. Um outro contributo é a abertura da onomasiologia ao estudo da escolha que os falantes têm que fazer de uma expressão particular entre outras disponíveis para nomear determinado conceito ou referente – a chamada onomasiologia *pragmática* (GRONDELAERS & GEERAERTS, 2003), que releva do plano do uso real das expressões linguísticas, distinta da onomasiologia *estrutural*, que releva do plano da estrutura e constitui a modalidade tradicional da onomasiologia. Um importante contributo descritivo e teórico sobre a variação lexical é o trabalho de Geeraerts, Grondelaers & Bakema (1994). Por tudo isto, é natural que o conhecimento adquirido sobre a variação lexical constitua uma espécie de modelo para o estudo cognitivo das outras formas de variação linguística.

Importa distinguir três tipos de variação lexical:

- variação *semasiológica*: diferentes sentidos e/ou (tipos de) referentes de um item lexical, isto é, polissemia e prototipicidade (ver SILVA, 2006c);
- variação *onomasiológica conceptual*: diferentes categorias conceptuais para nomear determinado conceito ou tipo de referente; por exemplo, escolhas taxionómicas entre *jogador* e *atacante* ou entre *casaco* e *blusão*;

- variação *onomasiológica formal*: diferentes termos para nomear uma mesma categoria conceptual, isto é, sinónimos denotacionais, como *atacante*, *avançado* e *dianteiro* ou *casaco* e *blazer*.

A estes três tipos pode adicionar-se um quarto: a variação *contextual* ou variação sociolinguística e estilística, pragmático-discursiva e diacrónica. Há aqui uma diferença: a variação contextual constitui uma dimensão adicional para as variações semasiológica e onomasiológica categorial, ao passo que é justamente essa variação contextual a que define a variação onomasiológica formal, isto é, as diferenças entre os sinónimos denotacionais.

É precisamente a variação onomasiológica formal e, portanto, aquela que tem a ver com as escolhas efectivas entre sinónimos denotacionais o objecto mais específico da *sociolexicologia* e da nossa investigação, que apresentaremos a seguir. Com efeito, 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. Acrescente-se, por último, que estes tipos de variação lexical, incluindo o que aqui é objecto de pesquisa (variação onomasiológica formal), se podem aplicar à variação gramatical.

## 5. Lexicologia quantitativa e sociolectometria

Para medir a distância entre duas variedades lectais – no nosso caso, para medir a convergência e a divergência lexical entre as duas variedades nacionais do Português –, são utilizados dois métodos quantitativos: a medida de uniformidade (U) entre as variedades e a medida do impacto de determinado traço (A), ligado a itens ou a conceitos lexicais, na uniformidade entre as variedades. Ambas as medidas foram desenvolvidas por Geeraerts, Grondelaers & Speelman (1999), no seu trabalho sociolexicológico sobre as duas variedades nacionais do Neerlandês. A medida U ou medida da “uniformidade linguística baseada em perfis” (GEERAERTS, 2001; SPEELMAN, GRONDELAERS & GEEARERTS, 2003) é a medida fundamental. Envolve duas noções específicas: *perfil* (*onomasiológico formal*) ou conjunto de sinónimos denotacionais usados para designar determinado conceito ou função, 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 percentual obtém-se somando as frequências relativas mais pequenas de cada termo alternativo: 30% de A e 40% de B.

Tecnicamente, a uniformidade de um conceito é calculada pela seguinte fórmula:

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

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. 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 Zs ponderados pela frequência relativa G de Z dentro do conjunto total de Zs. Obtém-se o coeficiente de ponderação  $G_{Zi}$  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  $Z_n$ .

Convergência e divergência entre duas variedades exprimem-se em aumento e diminuição de  $U/U'$ , respectivamente. Os cálculos ponderados ( $U'$  e, a seguir,  $A'$ ) são mais representativos, porque têm em conta a frequência relativa dos diferentes conceitos e termos. Uma diferença inferior a 5% é estatisticamente irrelevante.

A proporção de termos com determinado traço, ou medida A, é calculada pelas seguintes duas fórmulas:

$$A_{K,Z}(Y) = F_{z,y}(X_i) \cdot W_{x_i}(K)$$

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.

$$A'_K(Y) = A_{K,Zi}(Y) \cdot G_{Zi}(Y)$$

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  $Y$ .

## **6. Estudo de caso: convergência e divergência lexical entre o português europeu e o português brasileiro**

Passamos a apresentar, de modo bastante sumário, os dados e os resultados principais de três questões de investigação sobre as relações lexicais entre o português europeu e o português brasileiro (doravante, PE e PB):

- (1) saber se as duas variedades nacionais estão num processo de convergência ou de divergência lexical nos últimos 50 anos;
- (2) saber que influência têm factores linguísticos internos na tendência evolutiva global;
- (3) saber se a distância lexical entre registo padrão e registos subpadrão (intermédios) é maior numa variedade ou na outra.<sup>1</sup>

As duas primeiras questões – as mais importantes da presente investigação – implicam uma análise diacrónica, sendo externa a primeira e interna a segunda. A última é a questão sincrónica da estratificação lexical. O objecto de análise é a variação onomasiológica formal, isto é, aquela que, como vimos na secção anterior, envolve sinónimos denotacionais; e a base empírica consiste em largos milhares de observações do uso de sinónimos alternativos designando 43 conceitos nominais de dois campos lexicais – futebol e vestuário. Está ainda em análise um terceiro campo lexical – o dos termos de saúde. Este estudo apoia-se na concepção geral e nos métodos quantitativos da investigação de sociolexicologia cognitiva desenvolvida por Dirk Geeraerts e sua equipa para o Neerlandês da Holanda e da Bélgica (GEERAERTS, GRONDELAERS & SPEELMAN, 1999).

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

Os materiais de investigação foram extraídos manualmente de três fontes: (i) jornais de desporto e revistas de moda dos primeiros anos das décadas de 50, 70 e 90-2000; (ii) linguagem da Internet de conversação *on-line* de IRC ou *chats*; e (iii) etiquetas de roupas de lojas de vestuário. Os materiais de (i) permitirão responder à questão da convergência/divergência e os materiais de (ii) e (iii) à questão da estratificação. Os materiais do PB provêm dos dois centros urbanos mais importantes do país: São Paulo e Rio de Janeiro. O sub-*corpus* de futebol compreende textos de 8 jornais desportivos com a extensão de 2,7 milhões de palavras e textos de *chats* com a extensão de 15 milhões de palavras. O sub-*corpus* de vestuário contém textos de 24 revistas de moda com a extensão de 1,2 milhões de palavras e 1.300 imagens de etiquetas de lojas de vestuário. Estes dois sub-*corpora* constituem o *corpus* CONDIVport, actualmente com 4 milhões de palavras do registo formal (jornais e revistas) e 15 milhões de palavras do registo informal (*chats* e etiquetas). Está parcialmente 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 nos jornais de desporto e 143.946 observações nos *chats*, 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 nas revistas de moda e 3.240 observações nas etiquetas de lojas de vestuário. Para não inflacionar as diferenças, foram excluídos os termos marcadamente populares. Pelas limitações de espaço, apresentamos apenas os nomes dos conceitos ou perfis na designação utilizada no PE:

- Futebol: ‘árbitro’, ‘árbitro auxiliar’, ‘avançado’, ‘baliza’, ‘bola’, ‘defesa’, ‘equipa’, ‘extremo’, ‘falta’, ‘finta’, ‘fora-de-jogo’, ‘golo’, ‘grande penalidade’, ‘guarda-redes’, ‘jogada’, ‘jogo’, ‘médio’, ‘pontapé’, ‘pontapé de canto’, ‘pontapé livre’, ‘treinador’
- Vestuário: ‘blusa F’, ‘blusão M/F’, ‘calças M/F’, ‘calças curtas M/F’, ‘calças justas F’, ‘camisa M’, ‘camisola M/F’, ‘casaco F’, ‘casaco M’, ‘casaco curto F’, ‘casaco curto M’, ‘casaco de cerimónia M/F’, ‘casaco de malha M/F’, ‘casaco impermeável M/F’, ‘casaco quente (Inverno) M/F’, ‘conjunto F’, ‘fato M’, ‘jaqueta M/F’, ‘jeans M/F’, ‘saia F’, ‘t-shirt M/F’, ‘vestido F’

Apenas dois exemplos de perfis onomasiológicos:

- GUARDA-REDES: *arqueiro, goal-keeper, goleiro, golquíper, guarda-meta, guarda-redes, guarda-vala(s), guardião, keeper, porteiro, quíper, vigia*
- CASACO F: *“blazer”, blêizer, casaco, casaquinho/a, “manteau”, mantô, paletó, “paletot”*

A questão (1), anteriormente enunciada, subdivide-se nas seguintes três questões específicas:

- (1.1) convergência ou divergência entre o PE e o PB?
- (1.2) a tendência convergente/divergente ocorre dos dois lados ou apenas de um?
- (1.3) a uniformidade interna de cada variedade aumenta ou diminui?

A Figura 1 sintetiza os resultados percentuais de uniformidade ponderada  $U'$  dos 21 perfis de futebol e dos 22 perfis de vestuário do português de Portugal (P) e do Brasil (B) das décadas de 50, 70 e 90-2000: os números nas linhas horizontais indicam a percentagem de  $U'$  entre o PE e o PB nos anos 50, 70 e 2000, respondendo assim à questão (1.1); os números nas linhas verticais e diagonais indicam a percentagem de  $U'$  de período para período e entre diferentes períodos e respondem à questão (1.2); os números subpostos a cada variedade e período indicam a percentagem de  $U'$  dentro de cada variedade, respondendo à questão (1.3).

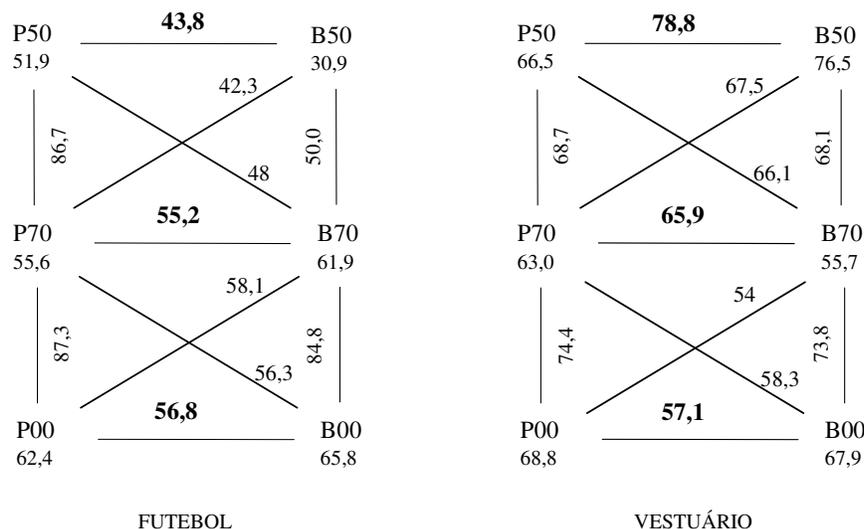


FIGURA 1: Resultados de uniformidade ponderada U' dos termos de futebol e vestuário

Em relação à questão (1.1) e no caso do futebol, as percentagens de U' das linhas horizontais indicam convergência de 50 para 70 e estabilidade a partir de então, mas as de U (que aqui não estão indicadas) não mostram qualquer alteração significativa: U' aumenta 11,4% entre 1950 e 1970 e somente 1,6% entre 1970 e 2000, ao passo que o valor U permanece mais ou menos o mesmo. Quer isto dizer que é em nível dos conceitos mais frequentes que a convergência se verifica. Ao mesmo tempo, os números mostram uma grande distância entre as duas variedades nos três períodos. Pelo contrário, no caso do vestuário as percentagens de U' evidenciam divergência ao longo do tempo: U' diminui 21,7% entre 1950 e 2000. Temos, portanto, resultados diferentes: convergência restringida no vocabulário do futebol e divergência clara no vocabulário do vestuário. A hipótese da divergência confirma-se no domínio do vestuário, mas não no domínio do futebol.

Para a questão (1.2) e no caso do futebol, as percentagens de U' das linhas verticais mostram que há mais mudanças do lado do PB. Além disso, de 50 para 70 o PB parece aproximar-se do PE (ver a linha diagonal): B70 está mais próximo, não somente de P70, mas também de P50 (48% contra 42,3%). 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 esta aparente aproximação. No caso do vestuário, as percentagens de U e U' das linhas verticais e diagonais indicam simetria nas tendências evolutivas das duas variedades e mostram que a divergência se dá dos dois lados nos dois períodos.

Para a questão (1.3), as percentagens de U' dentro de cada variedade e época mostram que há mais mudanças do lado do PB: em ambos os campos lexicais, o PB muda mais do que o PE. Além disso, as mudanças mais acentuadas dão-se de 50 para 70. No caso do futebol, há um grande aumento da uniformidade interna no PB, ao passo que as alterações no PE são menores. Isto poderá indicar maior estandardização do vocabulário do futebol no PB nos últimos anos e/ou menor preocupação com o apuramento estilístico da língua nos jornais brasileiros de hoje. No caso do vestuário, temos um padrão evolutivo idêntico: diminuição da uniformidade interna de 50 para 70 e aumento de 70 para 2000.

Passemos agora à questão (2), do impacto de factores linguísticos internos. Vamos medir o impacto de dois traços ligados aos itens lexicais em análise na tendência evolutiva global: 'endo/exogeneidade' e 'estrangeirismos'. No primeiro caso, queremos saber se existe uma orientação endógena ou exógena que faça com que uma das variedades se aproxime ou se afaste da outra. No segundo caso, queremos saber qual o impacto do inglês e do francês na tendência evolutiva global. Um terceiro factor, ainda em estudo, tem a ver com a influência normativa e por ele queremos saber qual a consequência da propaganda linguística na tendência evolutiva global.

A endo/exogeneidade deriva matematicamente do *corpus*: palavras *endógenas* são termos típicos de um sub-*corpus* (ocorrem muitas vezes nesse sub-*corpus*) e palavras *exógenas* são termos típicos do resto do *corpus*. Temos de calcular a proporção de palavras com os traços 'binacional' (AUNI), 'endógeno' (AENDO) e 'exógeno' (AEXO). Matematicamente, a evolução convergente implica aumento de AUNI e AEXO e diminuição de AENDO; o inverso determina a evolução divergente. Sem podermos apresentar aqui os respectivos cálculos, fiquemos apenas com a indicação dos resultados.

No futebol, temos aumento de A'UNI de 50 para 70, maior do lado do PB; diminuição de A'ENDO também de 50 para 70 e maior do lado do PB; e maior alteração de A'EXO no PB. Estes resultados condizem com a convergência

anteriormente observada (de 50 para 70) e confirmam que o PB muda mais do que o PE. O aumento de EXO é sempre maior do que a diminuição de ENDO tanto no PB como no PE, mas mais ainda naquele do que neste. Quer isto dizer que as duas variedades adotam mais itens exógenos do que abdicam de termos endógenos.

No vestuário, a divergência anteriormente verificada faz-se pela diminuição de A'UNI e, sobretudo, pelo aumento de A'ENDO. Nos anos 70 e nos anos 2000, surgem novos termos endógenos tanto no PB como no PE, o que explica a evolução divergente de ambos os lados. Esta endogeneidade nova tem a ver com a entrada no mundo da moda de termos novos ou de usos novos de termos já existentes.

Vejam agora a influência estrangeira. Vamos medir a proporção de itens com o traço (A) 'anglicismo', 'galicismo' ou 'estrangeirismo' (qualquer que seja a proveniência) no perfil onomasiológico de um conceito e, depois, no conjunto dos conceitos analisados nas amostras do PE e PB. A atribuição deste 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. As Tabelas 1 e 2 apresentam os resultados com as percentagens de anglicismos (A'<sub>Ingl</sub>), francesismos (A'<sub>Fr</sub>) e todos os estrangeirismos (A'<sub>estrangeir</sub>).

TABELA 1

Estrangeirismos no *corpus* do futebol

A' <sub>Ingl</sub> (P50)	7,1%	18,0%	A' <sub>Ingl</sub> (B50)
A' <sub>Ingl</sub> (P70)	9,8%	17,1%	A' <sub>Ingl</sub> (B70)
A' <sub>Ingl</sub> (P00)	10,2%	16,2%	A' <sub>Ingl</sub> (B00)
A' <sub>estrangeir</sub> (P50)	13,9%	23,5%	A' <sub>estrangeir</sub> (B50)
A' <sub>estrangeir</sub> (P70)	17,9%	22,8%	A' <sub>estrangeir</sub> (B70)
A' <sub>estrangeir</sub> (P00)	18,5%	23,3%	A' <sub>estrangeir</sub> (B00)

TABELA 2  
Estrangeirismos no *corpus* do vestuário

A' <sub>Fr</sub> (P50)	17,6%	18,5%	A' <sub>Fr</sub> (B50)
A' <sub>Fr</sub> (P70)	15,9%	18,1%	A' <sub>Fr</sub> (B70)
A' <sub>Fr</sub> (P00)	10,2%	7,9%	A' <sub>Fr</sub> (B00)
A' <sub>Ingl</sub> (P50)	3,3%	4,2%	A' <sub>Ingl</sub> (B50)
A' <sub>Ingl</sub> (P70)	5,8%	7,6%	A' <sub>Ingl</sub> (B70)
A' <sub>Ingl</sub> (P00)	16,9%	17,0%	A' <sub>Ingl</sub> (B00)
A' <sub>estrang</sub> (P50)	22,4%	23,8%	A' <sub>estrang</sub> (B50)
A' <sub>estrang</sub> (P70)	22,1%	26,7%	A' <sub>estrang</sub> (B70)
A' <sub>estrang</sub> (P00)	28,2%	24,9%	A' <sub>estrang</sub> (B00)

No *corpus* do futebol, a influência de anglicismos e de 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 maior 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 *corpus* do vestuário, verifica-se uma diminuição de francesismos (mais acentuada no PB) e um claro aumento de anglicismos, por razões socioculturais 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 continua a ser 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, a Tabela 3 mostra 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. Todos estes resultados explicam a emergência e também o tremendo fracasso das atitudes de purismo linguístico no Brasil.

TABELA 3

Adaptações/decalques de anglicismos no *corpus* de futebol

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)

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 crescente do PB sobre o PE, em particular no domínio do futebol, não se confirma claramente, como se pode verificar na Tabela 4. 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.

TABELA 4

Brasileirismos no *corpus* português de futebol

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%

Através do método estatístico da *análise de regressão*, pudemos verificar que a variável 'endogeneidade' é a que mais se correlaciona com a uniformidade U': esta correlação apresenta sempre resultados estatísticos significativos ( $p$  inferior a 0,05) e uma média de 99% da evolução de U' pode ser descrita com a fórmula A'ENDO. As variáveis 'anglicismos' e 'todos os 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 'endogeneidade' e, com exceção do futebol no PE, entre U' e 'anglicismos' e 'todos os estrangeirismos' (e também 'exogeneidade',

embora não significativa), isto é,  $U'$  diminui quando a proporção destas variáveis aumenta; e há uma correlação positiva entre  $U'$  e ‘francesismos’ no vocabulário do vestuário.

Finalmente a questão (3) da estratificação. Para calcular a distância entre o estrato padrão e o estrato subpadrão, vamos comparar, no caso do futebol, os dados do registo mais formal dos jornais dos anos 90/2000 (P00 e B00) e os dados do registo informal dos *chats* ( $P_{sub}$ 00 e  $B_{sub}$ 00) e, no caso do vestuário, 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 ( $P_{sub}$ 00 e  $B_{sub}$ 00). Os resultados são os seguintes:

Futebol:

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

Vestuário:

$$\begin{array}{lcl} U'(P00, P_{sub}00) & > & U'(B00, B_{sub}00) \\ 71\% & > & 57,9\% \end{array}$$

A hipótese de uma distância maior entre padrão e subpadrão no PB confirma-se no sub-*corpus* do vestuário, mas não no do futebol. No caso do futebol, há um grau relativamente elevado de uniformidade entre padrão (jornais) e subpadrão (*chats*) dentro de cada uma das variedades, o que poderá indicar padronização lexical na conversação pela Internet. Os utilizadores de *chat* tendem a ser conformistas quando se referem a conceitos tão populares como os do futebol. Além disso, a informalidade da linguagem do *chat* é determinada mais por outros factores linguísticos do que pelos factores lexicais. Já a uniformidade entre as duas variedades em nível do estrato subpadrão desce consideravelmente para um valor muito próximo do da uniformidade em nível do estrato padrão: 59,2% e 56,8%, respectivamente. Ou seja: a grande distância entre as duas variedades verifica-se em nível tanto do padrão como do subpadrão. No caso do vestuário, há ainda a verificar que a distância entre as duas variedades é maior em nível do estrato subpadrão (45,8%) do que em nível do estrato padrão (57,1%).

## 7. Extensões lexicais e gramaticais

A fase seguinte da investigação pretende estender o trabalho realizado em três direcções: (i) ainda no domínio lexical, vamos incluir palavras de outros campos lexicais, como os termos de saúde, e passar das palavras de conteúdo para as palavras funcionais, particularmente as preposições; (ii) em direcção ao domínio gramatical, pretendemos analisar parâmetros não-lexicais, designadamente variáveis sintácticas e morfológicas; (iii) ampliar e refinar a situação estratificacional de cada variedade, acrescentando mais variáveis estilísticas, em ordem ao estudo dos indicadores (lexicais e não-lexicais) de subpadronização.

Mais especificamente, pretendemos analisar as correlações entre variáveis lexicais e variáveis não-lexicais e comparar o seu impacto na convergência/divergência, na estratificação – particularmente as hipóteses de assimetria estratificacional (distância maior entre padrão e subpadrão no PB) e maiores alterações de uniformidade interna no PB (sinal de padronização tardia) – e na subpadronização das duas variedades nacionais. Será que palavras funcionais, como preposições, repetirão os resultados obtidos com os termos de futebol e vestuário? Será que as variáveis não-lexicais contribuirão de modo diferente das variáveis lexicais para a estrutura variacional? Quantos mais níveis de estratificação deverão ser tomados em conta? Será a estratificação multidimensional? Como determinar o impacto mútuo dos vários factores variacionais?

Relativamente à questão principal da convergência/divergência e tendo em conta a distinção entre léxico (mais consciência) e gramática (menos consciência), a hipótese a confirmar é a de maior divergência: as palavras funcionais e as construções gramaticais divergirão mais do que as palavras de conteúdo.

Serão analisados 15 perfis de preposições e 20 perfis de construções sintácticas, seguindo os mesmos métodos onomasiológico e de uniformidade baseada em perfis (a ‘sinonímia’ de preposições e construções sintácticas é entendida no sentido de equivalência funcional). Os perfis preposicionais são restringidos aos mesmos tipos de complementos e ao mesmo contexto sintagmático, de forma a satisfazer a condição de sinonímia denotacional (p.ex. *reflectir/pensar em/sobre, interesse em/por, por/através de/via fax/internet/correio*). Os perfis construcionais são seleccionados de entre construções causativas/perceptivas (*fazer/mandar/deixar/ver y + INF flex/não-flex / INF não-flex + y / oração de que*), construções causativas sintéticas *vs.* analíticas, construções completivas e adverbiais finitas *vs.* infinitivas, construções de elevação,

construções de sujeito nulo e de objecto nulo, construções alternativas de indicativo/conjuntivo, padrões alternativos de ordem de palavras, colocações alternativas do adjectivo atributivo e, eventualmente, outras construções. Analisaremos, ainda, formações diminutivas e aumentativas, sufixos derivacionais alternativos, formas alternativas dos pronomes possessivos e pessoais, verbos de participio duplo e formações de plural. As técnicas de regressão linear e multilinear permitirão analisar as correlações entre estas variáveis dependentes e entre elas e as variáveis independentes estilísticas (jornais de qualidade > jornais populares > Internet *off-line* > Internet *on-line*, oralidade).

## 8. Conclusão

O estudo cognitivo das línguas, representado pela Linguística Cognitiva, encerra actualmente duas tensões correlacionadas: uma temática, entre o individual-neurofisiológico-universal e o interindividual-sociocultural-variacional; outra metodológica, entre introspecção e empiricismo. É um mito acreditar que a linguagem é geneticamente constringida a ser uniforme ou que as comunidades linguísticas são homogéneas. E é um equívoco pensar que fenómenos não-objectivistas como os cognitivos e os linguísticos não podem ser objectivamente estudados. E se a Linguística Cognitiva quer ser, de facto, um *modelo baseado no uso*, então tem que resolver aquelas duas tensões explorando o seu segundo elemento. Esperamos ter conseguido mostrar que a perspectiva cognitiva da linguagem implica a integração sistemática da variação social e da metodologia empírica baseada nos dados de um *corpus* (ou nos dados experimentais) e em técnicas quantitativas e estatísticas para os analisar. Estes dois desenvolvimentos implicam-se mutuamente e cada um deles tem as suas exigências. A integração da variação social implica a adopção da metodologia do *corpus* e exige a reinterpretção da cognição como *cognição social*. A adopção da metodologia do *corpus* exige sofisticação em técnicas quantitativas e estatísticas e implica um ciclo empírico-interpretativo de refinamento gradual das hipóteses pelo confronto repetido com os dados empíricos. A convergência destes dois desenvolvimentos marca uma das principais extensões da Linguística Cognitiva nestes primeiros anos do séc. XXI – o desenvolvimento da emergente Sociolinguística Cognitiva.

Esperamos ter conseguido mostrar também as vantagens da perspectiva cognitiva, do método da variação onomasiológica entre sinónimos denotacionais

e de métodos quantitativos multivariacionais no estudo da convergência e divergência entre o português europeu e o português brasileiro. Os resultados da investigação sociolxicológica realizada mostram que a hipótese da divergência se confirma 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 padronização do vocabulário do futebol. Além disso, não parece haver uma orientação de uma variedade em relação à outra; a variedade brasileira muda mais do que a europeia, é mais permeável à importação e adaptação de estrangeirismos e apresenta uma distância maior entre estrato padrão e subpadrão. Precisamos de consolidar a análise sociolxicológica, incluindo outros campos lexicais, e precisamos de a estender ao domínio das palavras funcionais e das variáveis não-lexicais. A análise das correlações entre variáveis lexicais e não-lexicais e entre estas variáveis dependentes e as variáveis independentes estilísticas permitirá um conhecimento mais exacto da verdadeira situação actual de divergência ou convergência entre as duas variedades nacionais e da estrutura estratificacional de cada uma. Com a sociolinguística cognitiva e quantitativa, pretendemos contribuir para o desenvolvimento da sociolinguística e da sociolectometria do português europeu e brasileiro e estamos certos de que os resultados obtidos garantirão uma base empírica segura para a construção de melhores políticas da língua portuguesa, como língua plurinacional e transcontinental.

## Nota

<sup>1</sup> Este estudo faz parte do projecto de investigação “Convergência e Divergência no Léxico do Português”, financiado pela Fundação para a Ciência e a Tecnologia de Portugal (Ref<sup>a</sup> POCTI/LIN/ 48575/2002), no âmbito do POCTI do fundo comunitário europeu FEDER. Para descrições parciais mais desenvolvidas dos resultados deste projecto, ver Silva & Duarte (2005) e Silva (2006a, b, no prelo).

## Referências Bibliográficas

- BARLOW, M.; KEMMER, S. (Eds.). *Usage-based Models of Language*. Stanford: CSLI Publications, 2000.
- BERNÁRDEZ, E. Intimate enemies? On the relations between language and culture. In: SILVA, A. S. da, TORRES, A.; GONÇALVES, M. (Org.). *Linguagem, cultura e cognição: estudos de Linguística Cognitiva*. Coimbra: Almedina, 2004. V. I, p. 21-45.
- BERNÁRDEZ, E. Social cognition: variation, language, and culture in a cognitive linguistic typology. In: RUIZ DE MENDOZA, F.J.; PEÑA CERVEL, S. (Ed.). *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interaction*. Berlin/New York: Mouton de Gruyter, 2005. p. 191-222.
- BLANK, A.; KOCH, P. (Ed.). *Historical Semantics and Cognition*. Berlin/New York: Mouton de Gruyter, 1999.
- BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.
- BYBEE, J.; HOPPER, P. (Ed.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 2001.
- CROFT, W.; CRUSE, D. A. *Cognitive Linguistics*. Cambridge: Cambridge University Press, 2004.
- DAMÁSIO, A. *O erro de Descartes*. Emoção, razão e cérebro humano. Mem Martins: Publicações Europa-América, 1995.
- DAMÁSIO, A. *O sentimento de si*. O corpo, a emoção e a neurobiologia da consciência. Mem Martins: Publicações Europa-América, 2000.
- DIRVEN, R.; FRANK, R.; ILIE, C. (Ed.). *Language and Ideology*. Vol. II: Descriptive cognitive approaches. Amsterdam: John Benjamins, 2001.
- DIRVEN, R.; FRANK, R.; PÜTZ, M. (Ed.). *Cognitive Models in Language and Thought*. Ideology, Metaphors and Meanings. Berlin/New York: Mouton de Gruyter, 2003.
- DIRVEN, R.; HAWKINS, B.; SANDIKCIOGLU (Ed.). *Language and Ideology*. Vol. I: Theoretical cognitive approaches. Amsterdam: John Benjamins, 2001.
- EDELMAN, G. M. *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books, 1992.
- FAUCONNIER, G. *Mappings in Thought and Language*. Cambridge: Cambridge University Press, 1997.

FAUCONNIER, G.; TURNER, M. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books, 2002.

FRANK, R. M.; DIRVEN, R.; ZIEMKE, T.; BERNÁRDEZ, E. (Ed.). *Body, Language, and Mind II: Sociocultural Situatedness*. Berlin: Mouton de Gruyter, 2008.

GEERAERTS, D. *Diachronic Prototype Semantics. A Contribution to Historical Lexicology*. Oxford: Clarendon Press, 1997.

GEERAERTS, D. On measuring lexical convergence. In: SILVA, A. S. (Org.). *Linguagem e cognição: a perspectiva da Linguística Cognitiva*. Braga: Associação Portuguesa de Linguística / Universidade Católica Portuguesa, 2001. p. 51-61.

GEERAERTS, D. Cultural models of linguistic standardization. In: DIRVEN, R.; FRANK, R.; PÜTZ, M. (Ed.). *Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings*. Berlin/New York: Mouton de Gruyter, 2003. p. 25-68. (republicado em SILVA, A. S. da; TORRES, A.; GONÇALVES, M. (Org.). *Linguagem, cultura e cognição: Estudos de Linguística Cognitiva*. Coimbra: Almedina, 2004. V. I. p. 47-84.)

GEERAERTS, D. Lectal data and empirical variation in Cognitive Linguistics. In: RUIZ DE MENDOZA, F. J.; PEÑA CERVEL, S. (Ed.). *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interactions*. Berlin/New York: Mouton de Gruyter, 2005. p. 163-189.

GEERAERTS, D. Methodology in Cognitive Linguistics. In: KRISTIENSEN, G.; ACHARD, M.; DIRVEN, R.; RUIZ DE MENDOZA, F. J. (Ed.). *Cognitive Linguistics: Current Applications and Future Perspectives*. Berlin/New York: Mouton de Gruyter, 2006. p. 21-49.

GEERAERTS, D. (Ed.). *Cognitive Linguistics: Basic Readings*. Berlin/New York: Mouton de Gruyter, 2006.

GEERAERTS, D.; GRONDELAERS, S.; BAKEMA, P. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter, 1994.

GEERAERTS, D.; GRONDELAERS, S.; SPEELMAN, D. *Convergentie en Divergentie in de Nederlandse Woordenschat*. Amsterdam: Meertens Instituut, 1999.

GIBBS, R. W. *The Poetics of Mind. Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press, 1994.

GIBBS, R. W. *Embodiment and Cognitive Science*. New York: Cambridge University Press, 2006.

GONZALEZ-MARQUEZ, M.; MITTELBERG, I.; COULSON, S.; SPIVEY, M. J. (Ed.). *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, 2007.

GRIES, S. T. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London: Continuum Press, 2003.

GRIES, S. T.; STEFANOWITSCH, A. (Ed.). *Corpora in Cognitive Linguistics*. Corpus-based Approaches to Syntax and Lexis. Berlin/New York: Mouton de Gruyter, 2006.

GRONDELAERS, S.; GEERAERTS, D. Towards a pragmatic model of cognitive onomasiology. In: CUYCKENS, H.; DIRVEN, R.; TAYLOR, J. (Ed.). *Cognitive Approaches to Lexical Semantics*. Berlin/New York: Mouton de Gruyter, 2003. p. 67-92.

GRONDELAERS, S.; GEERAERTS, D.; SPEELMAN, D. A case for a Cognitive corpus linguistics. In: GONZALEZ-MARQUEZ, M.; MITTELBERG, I.; COULSON, S.; SPIVEY, M. J. (Ed.). *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, 2007. p. 149-169.

GRONDELAERS, S.; SPEELMAN, D.; GEERAERTS, D. Regressing on *er*. Statistical analysis of texts and language variation. In: MORIN, A.; SÉBILLOT, P. (Ed.). *6<sup>th</sup> International Conference on the Statistical Analysis of Textual Data*. Rennes: Institut National de Recherche en Informatique et en Automatique, 2002. p. 335-346.

HARDER, P. The status of linguistic facts. Rethinking the relation between cognition, social institution, and utterance from a functional point of view. *Mind and Language* 18, p. 52-76, 2003.

HOLLAND, D.; QUINN, N. (Eds.). *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press, 1987.

HOPPER, P. J.; TRAUOGOTT, E. C. *Grammaticalization*. Cambridge: Cambridge University Press, 1993.

KÖVECSES, Z. *Metaphor and Emotion*. Language, Culture, and Body in Human Feeling. Cambridge-Paris: Cambridge University Press, 2000.

KÖVECSES, Z. *Metaphor in Culture: Universality and Variation*. Cambridge: Cambridge University Press, 2005.

KRISTIANSEN, G. How to do things with allophones: Linguistic stereotypes as cognitive reference points in social cognition. In: DIRVEN, R.; FRANK, R.; PÜTZ, M. (Ed.). *Cognitive Models in Language and Thought: Ideologies, Metaphors, and Meanings*. Berlin/New York: Mouton de Gruyter, 2003. p. 69-120.

KRISTIANSEN, G.; DIRVEN, R. (Ed.). *Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter. (No prelo).

LAKOFF, G. The contemporary theory of metaphor. In: ORTONY, A. (Ed.). *Metaphor and Thought*, Cambridge: Cambridge University Press, 1993. p. 202-251.

LAKOFF, G. *Moral Politics: What Conservatives Know that Liberals Don't*. Chicago: The University of Chicago Press, 1996.

LAKOFF, G. The brain's concepts. The role of the sensory-motor system in reason and language. *Working papers of the NTL Group at the International Computer Science Institute (ICSI)*. Berkeley: University of California, Berkeley, 2003.

LAKOFF, G.; JOHNSON, M. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books, 1999.

LANGACKER, R. W. *Foundations of Cognitive Grammar*. V. 1: Theoretical Prerequisites. Stanford: Stanford University Press, 1987.

LANGACKER, R. W. A usage-based model. In: RUDZKA-OSTYN, B. (Ed.). *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 1988. p.127-161.

LANGACKER, R. W. *Foundations of Cognitive Grammar*. V. 2: Descriptive Application. Stanford: Stanford University Press, 1991.

LANGACKER, R. W. A dynamic usage-based model. In: BARLOW, M.; KEMMER, S. (Ed.). *Usage-based Models of Language*. Stanford: CSLI Publications, 2000. p. 1-63.

LEVINSON, S. C. *Space in Language and Cognition*. Cambridge: Cambridge University Press, 2003.

PALMER, G. B. *Toward a Theory of Cultural Linguistics*. Austin: University of Texas Press, 1996.

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

SILVA, A. S. Sociolinguística cognitiva e o estudo da convergência/divergência entre o Português Europeu e o Português Brasileiro. *Veredas 10. Revista de Estudos Lingüísticos*. Disponível em: <<http://www.revistaveredas.ufjf.br>>. 2006b.

SILVA, A. S. *O mundo dos sentidos em português: polissemia, semântica e cognição*. Coimbra: Almedina, 2006c.

SILVA, A. S. Medindo a convergência e divergência lexical entre o Português Europeu e o Português Brasileiro. In: *Anais do II Congresso da Associação Internacional de Linguística do Português*, Universidade Federal do Rio de Janeiro, set. 2007. p. 25-27. (No prelo).

SILVA, A. S.; DUARTE, M. O léxico do vestuário no Português Europeu e no Português Brasileiro: convergência ou divergência? *Revista Portuguesa de Humanidades* 9, p. 117-136, 2005.

SINHA, C. Grounding, mapping, and acts of meaning. In: JANSSEN, T.; REDEKER, G. (Ed.). *Cognitive Linguistics: Foundations, Scope, and Methodology*. Berlin/New York: Mouton de Gruyter, 1999. p. 223-255.

SINHA, C.; JENSEN DE LÓPEZ, K. Language, culture, and the embodiment of spatial cognition. *Cognitive Linguistics* 11, p. 17-41, 2000

SPEELMAN, D.; GRONDELAERS, S.; GEERAERTS, D. Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37, p. 317-337, 2003.

STEFANOWITSCH, A. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In: ROHDENBURG, G.; MONDORF, B. (Ed.). *Determinants of Grammatical Variation*. Berlin/New York: Mouton de Gruyter, 2003. p. 413-444.

STEFANOWITSCH, A.; GRIES, S. T. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8, p. 209-243, 2003.

STEFANOWITSCH, A.; GRIES, S. T. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1, p. 1-43, 2005.

TALMY, L. *Toward a Cognitive Semantics*. Cambridge: The MIT Press, 2000. 2 v.

TALMY, L. Foreword. In: GONZALEZ-MARQUEZ, M.; MITTELBERG, I.; COULSON, S.; SPIVEY, M. J. (Ed.). *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, 2007. p. XI-XXI.

TOMASELLO, M. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press, 1999.

TOMASELLO, M. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11, p. 61-82, 2000.

TOMASELLO, M. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press, 2003.

TRAUGOTT, E. C.; DASHER, R. B. *Regularity in Semantic Change*. Cambridge: Cambridge University Press, 2002.

TUMMERS, J.; HEYLEN, K.; GEERAERTS, D. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1, p. 225-261, 2005.

VARELA, F.; THOMPSON, E.; ROSCH, E. *The Embodied Mind*. Mass.: Massachusetts Institute of Technology, 1991.

ZIEMKE, T.; ZLATEV, J.; FRANK, R. M. (Ed.). *Body, Language, and Mind I: Embodiment*. Berlin/New York: Mouton de Gruyter. (No prelo).

ZLATEV, J. *Situated Embodiment: Studies in the Emergence of Spatial Meaning*. Stockholm: Gotab, 1997.

ZLATEV, J. Situated embodied semantics and connectionist modeling. In: ALLWOOD, J.; GARDENFORS, P. (Ed.). *Cognitive Semantics: Meaning and Cognition*. Amsterdam: John Benjamins, 1999. p. 173-194.

ZLATEV, J. Beyond cognitive determination. Interactionism in the acquisition of spatial semantics. In: LEATHER, J.; VAM DAM, J. (Ed.). *Ecology of Language Acquisition*. Amsterdam: Kluwer Academic Publishers, 2003. p. 83-107.