

## **SISTEMA E VARIAÇÃO: QUÃO SISTEMÁTICO PODE SER O SISTEMA LINGUÍSTICO NUM MODELO BASEADO NO USO?**

*Augusto Soares da Silva (UCP)\**

### **RESUMO**

Argumentaremos que um modelo baseado no uso implica uma conceção de língua como sistema dinâmico complexo, que correlaciona os aspetos sociais e os aspetos cognitivos. Descritivamente, um modelo baseado no uso implica a inclusão da variação intralinguística no estudo de qualquer expressão e a sua correlação com o significado conceptual. Metodologicamente, um modelo baseado no uso implica a implementação de métodos empíricos multivariacionais, como os métodos socioletométricos. No plano metateórico, um modelo baseado no uso implica uma des-sistematização do sistema linguístico e a construção de um modelo multifatorial de gramática. Mostraremos como a Sociolinguística Cognitiva, uma extensão emergente da Linguística Cognitiva como modelo orientado para o significado e para o uso, consegue dar conta destas implicações. Como ilustração, apresentaremos os resultados da nossa investigação sociocognitiva e socioletométrica sobre a divergência entre o português europeu e brasileiro.

**PALAVRAS-CHAVE:** modelo baseado no uso, sistema dinâmico complexo, Sociolinguística Cognitiva, variação linguística, português europeu e brasileiro

### **1. INTRODUÇÃO <sup>1</sup>**

Numa conceção de linguagem baseada no uso efetivo das línguas ou, usando o termo introduzido por Saussure, numa conceção de linguagem no plano da parole, os sistemas linguísticos são, inevitavelmente, não-homogéneos. A razão é simples: a variação linguística é a consequência imediata e inevitável do uso da língua, pelo que uma comunidade linguística nunca pode ser totalmente homogénea; a variação social é intrínseca a qualquer comunidade e, logo, a qualquer comunidade linguística. Um modelo linguístico baseado no uso da língua tem pois de abandonar a ideia da homogeneidade linguística, presente na conceção idealista tradicional da linguagem, tanto da linguística estrutural de Saussure como, e sobretudo, da linguística gerativa de Chomsky. Lembremos o postulado de Chomsky do “falante/ouvinte ideal, que conhece perfeitamente a sua língua”. A ideia de uma comunidade linguística homogénea é, de facto, inverosímil e mesmo mítica.

---

\* Professor da Universidade Católica Portuguesa – Braga, Portugal. E-mail: [assilva@braga.ucp.pt](mailto:assilva@braga.ucp.pt)

1. Este trabalho foi financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia de Portugal no âmbito do projeto de investigação “PEst-OE/FIL/UI0683/2011”.

Mas o que é o sistema de uma língua num modelo linguístico baseado no uso? Sendo o uso de uma língua inevitavelmente variável, quão (a)sistemático é o sistema de uma língua? Como é possível definir uma língua se o seu uso é intrinsecamente heterogéneo? Mais ainda, sendo a variação no uso letalmente estruturada, isto é, organizada em termos de dialetos, socioletos e idioletos ou registos, quão sistemática é essa variação?

Neste estudo, procuraremos responder à pergunta em epígrafe com a noção de sistema dinâmico complexo e mostrando que a variação letal (termo genérico para designar qualquer tipo de variação de uma língua: dialetos, variedades nacionais, socioletos, registos, estilos, idioletos) é uma parte integrante de um modelo linguístico baseado no uso. Tanto a língua como sistema dinâmico complexo como a sua inevitável variação letal integram e correlacionam fatores sociais e fatores cognitivos, justamente porque uma língua é tanto uma realidade social quanto uma realidade cognitiva. Um modelo linguístico baseado no uso deverá dar conta da interação dos aspetos cognitivos e dos aspetos sociais dos fenómenos linguísticos. Referiremos a Sociolinguística Cognitiva (Kristiansen & Dirven 2008, Geeraerts, Kristiansen & Peirsman 2010) – extensão emergente da Linguística Cognitiva (Geeraerts & Cuyckens 2007) como modelo baseado no uso e orientado para o significado –, como resposta teórica, descritiva e metodológica à visão da língua como sistema dinâmico complexo e à análise da interação entre os aspetos sociais e os aspetos cognitivos da variação letal. Entre os métodos que permitem analisar um sistema dinâmico complexo na sua relação dialética entre sistema e uso e na sua inevitável multivariacionalidade, destacaremos a socioletometria, cujos métodos quantitativos e multivariacionais conseguem medir distâncias entre variedades letais e correlacioná-las com todos os tipos de variáveis sociolinguísticas. Como ilustração empírica da visão sociocognitiva da língua como sistema dinâmico complexo e da abordagem socioletométrica da variação letal, apresentaremos, de forma muito sumária, devido às limitações de espaço do presente artigo, os elementos e os resultados da nossa investigação sobre a divergência entre o português europeu e o português brasileiro nos últimos 60 anos.

## **2. SISTEMA DINÂMICO COMPLEXO**

O facto primário do estudo linguístico é o comportamento linguístico dos falantes, isto é, a sua interação verbal. São os falantes individuais em interação uns com os outros o que o linguista pode observar e pode diretamente descrever. Naturalmente que somos levados a pensar que há alguma estrutura subjacente ao comportamento linguístico individual dos falantes, quer a nível individual, no sentido de que cada indivíduo tem o seu próprio sistema linguístico interno, isto é, o seu próprio conhecimento da língua, quer a nível interindividual, como algo partilhado pelos diferentes falantes de uma comunidade. Mas o que habitualmente pensamos e designamos como sistema linguístico é algo de abstrato, que o linguista infere a partir da observação do uso da língua; por outras palavras, o sistema é uma abstração secundária, que pretende dar conta das regularidades no uso da língua. Tanto para os linguistas, que procuram descrever o sistema de uma língua, como para os falantes, que procuram aprender o sistema de uma língua, o sistema linguístico não é diretamente acessível, mas apenas adquirido por abstração.

Como é então possível que os falantes aprendam, como crianças ou como adultos, o sistema de uma língua se o sistema não é diretamente acessível? Os falantes aprendem o sistema linguístico por alinhamento mútuo uns com os outros, isto é, eles aprendem por imitação e adaptação, necessariamente a nível interacional. O alinhamento linguístico dos falantes é incompleto, na medida em que não é possível interagir com todos os falantes de uma língua, mas apenas com uma pequena parte. É assim

que as interações comunicativas não só são eventos sociais como também refletem uma estrutura social, formada por grupos e redes sociais a que os falantes individuais pertencem e que configuram essas interações verbais.

Consequentemente, a variação letal (dialetal, socioletal, idioletal) é uma parte integrante da conceção do sistema linguístico baseada no uso (cf. Geeraerts 2010a). Uma língua é, afinal, um cluster de diferentes variedades letais, um cluster de subsistemas na conceção abstrata da língua. Até aqui, abandonamos a ideia saussureana e chomskyana de língua como sistema homogéneo, substituindo-a pela ideia de sistema linguístico heterogéneo, poliletal ou multiletal, no sentido da sociolinguística variacionista de Labov (1972). Mas a assunção de um modelo linguístico baseado no uso deve levar-nos a uma visão mais dinâmica da língua, a uma ideia de língua como sistema heterogéneo num sentido mais radical.

Sendo uma língua uma realidade não-homogénea, serão os seus dialetos, socioletos e idioletos, num sentido mais geral, os seus letos internamente homogéneos? Não são. Como já Chambers & Trudgill (1998) observaram, as variedades linguísticas fazem parte de um continuum dialetal, em que as variedades adjacentes são mutuamente compreensíveis, mas as variantes dos extremos da cadeia podem não o ser. Usando a terminologia da Linguística Cognitiva (Geeraerts & Cuyckens 2007), em particular a teoria do protótipo como modelo de categorização linguística e concetual (Taylor 1995, Geeraerts 1997), as variedades letais constituem categorias estruturadas com base em protótipos, isto é, algumas realizações são mais típicas, mais salientes, “melhores exemplares” de determinada variedade linguística do que outras realizações. Consequentemente, determinadas realizações, determinadas variáveis são mais centrais ou mais próximas do centro prototípico e outras mais afastadas, podendo estas últimas entrar em áreas de interseção com realizações ou variáveis de outras variedades linguísticas. Por outras palavras, cada variedade letal apresenta uma estrutura radial a partir de um centro prototípico.

Neste sentido mais radical de heterogeneidade linguística, cada língua é um sistema dinâmico complexo e, neste sistema, cada expressão, cada construção, cada palavra, cada forma linguística tem a sua própria distribuição letal. Esta distribuição letal é inevitavelmente multifatorial, já que envolve não apenas diferenças geográficas, mas também muitos outros fatores letais, sejam sociais, sejam estilísticos, sejam ainda características dos falantes. Mais ainda, estes diversos fatores letais interagem com os fatores conceptuais.

A necessária desconstrução da noção de sistema linguístico ou, por outras palavras, a necessária des-sistematização do sistema linguístico envolve dois argumentos essenciais: se “a língua” existe como sistema, então só pode existir como sistema dinâmico complexo; e se as variedades letais existem como entidades separadas dentro desse sistema dinâmico complexo, então elas só podem existir como categorias prototípicas, isto é, com uma estrutura interna assente em protótipos. A maior consequência descritiva e metodológica desta des-sistematização do sistema linguístico é a inclusão sistemática da variação letal na descrição de qualquer expressão linguística e a utilização de métodos multivariacionais capazes de darem conta da heterogeneidade e da multifatorialidade dos fenómenos linguísticos.

### 3. A ESTRUTURA DA VARIAÇÃO LINGUÍSTICA E TIPOS DE VARIAÇÃO DE SIGNIFICADO

Vamos ver como a variação linguística está estruturada e, ao mesmo tempo, como essa variação afeta a variação do significado. Consideremos o caso do significado lexical, que tomaremos como modelo de outros tipos de variação linguística.

As escolhas lexicais que os falantes fazem no discurso são determinadas por diferentes fatores. Obviamente que há escolhas lexicais de conceitos (mais ou menos) específicos determinadas pelo tema do discurso, mas há outras escolhas lexicais que têm a ver, não com diferenças entre conceitos, mas com diferenças sociolinguísticas, estilísticas ou pragmáticas. Por exemplo, a escolha entre guarda-redes e goleiro é uma escolha de formas que exprimem o mesmo conceito mas pertencem a diferentes variedades nacionais (guarda-redes é usado no português europeu, ao passo que goleiro se usa no português brasileiro); a escolha entre morrer e falecer é uma escolha de formas que exprimem o mesmo conceito mas são estilisticamente diferenciadas (falecer é usado em registos formais); a escolha entre você e o senhor é uma escolha de formas que exprimem o mesmo conceito mas são pragmaticamente diferenciadas. Podemos designar esta variação entre sinónimos denotacionais como variação onomasiológica formal, em oposição à variação onomasiológica conceptual, que envolve diferenças conceptuais, como entre guarda-redes/goleiro e jogador. Distinta da variação onomasiológica entre diferentes expressões alternativas para designar determinado conceito ou função, é a variação semasiológica, que envolve a escolha entre diferentes sentidos e/ou referentes de determinada palavra ou outra expressão<sup>2</sup>. Um quarto tipo de variação lexical é a variação contextual e tem a ver com todos os aspetos da situação comunicativa, não apenas as características mais ou menos permanentes dos falantes (como ser português ou brasileiro), mas também as características mais transitórias e interacionais do ato de fala (como o género de discurso, por exemplo).

Aos três primeiros tipos de variação lexical estão associados diferentes fenómenos de saliência lexical. A saliência semasiológica envolve os fenómenos da prototipicidade (Taylor 1995, Geeraerts 1997) e da polissemia (Silva 2006), isto é, o facto de alguns referentes ou alguns sentidos de determinada palavra serem mais representativos, em termos de frequência ou de coerência semântica, do que outros. Por exemplo, a maçã, a laranja ou a pera são referentes mais salientes da palavra fruto e (mais diretamente) fruta do que o limão, a romã ou o coco; e os sentidos de ‘abandonar’ e ‘permitir’ são mais salientes do verbo deixar do que o sentido de ‘ir-se embora, retirar-se’. A saliência na variação onomasiológica conceptual tem a ver com o chamado nível básico das hierarquias lexicais e com o incrustamento ou familiarização de determinados itens lexicais em relação a outros do mesmo paradigma. Por exemplo, saia e calças são categorias de nível básico, mais salientes do que minissaia e jeans. A saliência na variação onomasiológica formal consiste na prevalência sociolinguística entre sinónimos denotacionais, isto é, determinados termos são geográfica, social ou estilisticamente mais salientes do que os seus sinónimos. Crucialmente, estudar a saliência lexical é estudar a interação entre estrutura e uso, já que a saliência é a manifestação do uso na estrutura – algumas partes da estrutura são mais importantes do que outras, justamente porque são mais usadas.

---

\* A semasiologia toma como ponto de partida a palavra ou outra expressão para analisar os diferentes sentidos nela associados, ao passo que a onomasiologia toma o conceito ou função como ponto de partida e investiga as diferentes palavras ou outras expressões que o/a designam. Esta distinção remonta ao estudo seminal de Baldinger (1964). Sobre esta distinção e sua importância, ver também Silva (2006, 2011b). Sobre tipos de variação lexical, ver o estudo empírico de Geeraerts, Grondelaers & Bakema (1994), a quem se deve a distinção entre variação onomasiológica conceptual e variação onomasiológica formal.

Um outro facto crucial é que os quatro tipos de variação lexical se podem combinar de diferentes modos. Por exemplo, a variação entre deixar e abandonar (cf. Silva 1999) é conceptual, pelas diferenças de intensidade do abandono (maior em abandonar do que em deixar), e é também formal, pelas diferenças emotivas e estilísticas (abandonar é mais emotivo e estilisticamente mais marcado do que deixar). Por outro lado, variação onomasiológica e variação semasiológica podem condicionar-se mutuamente: por exemplo, a entrada dos verbos permitir e abandonar nos finais do português antigo foi um dos principais fatores para sucessivas reestruturações dos centros prototípicos de deixar e essas reestruturações prototípicas contribuíram para a diferenciação dos três verbos (Silva 1999). A escolha de um termo (in)formal pode correlacionar-se com fatores conceptuais: por exemplo, utilizamos mais palavras informais para temas familiares do que para temas científicos. A variação letal pode correlacionar-se com fatores conceptuais: por exemplo, conceitos menos frequentes, vagos, novos ou negativamente emotivos tendem a ser expressos por mais sinónimos dialetais (Geeraerts & Speelman 2010). E as escolhas conceptuais podem ser determinadas também por fatores sociais.

O que acontece com o significado lexical pode também verificar-se com o significado gramatical. Na variação construcional, pode distinguir-se a variação onomasiológica conceptual (construções alternativas exprimem diferentes funções), a variação onomasiológica formal (construções alternativas exprimem a mesma função) e a variação contextual. Fatores conceptuais, formais ou contextuais podem determinar as escolhas entre construções. Por exemplo, a escolha da construção completiva infinitiva mono-oracional (em detrimento da bioracional) selecionada por verbos causativos e percetivos (dita “união de orações”) é determinada por fatores conceptuais e/ou letais (a construção mono-oracional é mais produtiva na variedade europeia do que na brasileira – ver Silva 2005a).

As escolhas lexicais e gramaticais fazem-se, pois, em função de um ou mais do que um de três fatores: significado, forma e contexto. O mesmo é dizer que significado, forma e contexto são três fontes de variação linguística. Importa sublinhar que a variação onomasiológica formal, tanto lexical como construcional, é uma variável sociolinguística, permeável, como qualquer variável sociolinguística, a um conjunto de variáveis extralinguísticas, como a classe social, a idade, o sexo, a origem geográfica, o grupo social, o registo, etc. Os sinónimos denotacionais, lexicais ou funcionais, são interessantes do ponto de vista sociolinguístico, na medida em que evidenciam diferenças regionais, sociais, estilísticas e pragmáticas e são estas diferenças que motivam a existência e competição de variedades de uma língua.

Um outro problema semântico envolvido na variação linguística é a equivalência de significado, pré-requisito da noção de variável sociolinguística<sup>3</sup>. Até que ponto a noção de variável sociolinguística, com origem na fonologia, se pode aplicar ao léxico e à gramática? Como estabelecer equivalência semântica entre itens lexicais e entre construções morfológicas ou sintáticas? Em relação a itens lexicais como os de vestuário ou futebol, podemos controlar os seus referentes concretos. Isso é mais difícil em relação a funções linguísticas. Mesmo restringindo as variantes construcionais ao mesmo contexto sintagmático, a equivalência funcional completa não se dá nas línguas naturais: há sempre diferenças de perspetivação conceptual, como o evidenciam os modelos gramaticais cognitivos de Langacker (1987, 1991, 2008) e de Talmy (2000). Uma tentativa de resposta ao problema da equivalência semântica e à noção de variável sociolinguística, mais difíceis de resolver no domínio gramatical, é a seguinte: se as diferenças semânticas entre as variantes construcionais forem estáveis entre variedades linguísticas, então qualquer variação remanescente é variação sociolinguística.

---

3. Ver o debate inicial entre Lavandera (1978) e Labov (1978). Lavandera (1978: 171) argumenta que a aplicação da noção de variável sociolinguística aos domínios do léxico, morfologia e sintaxe requer uma “articulated theory of meanings”.

Szmrecsanyi (2010) e Coleman (2010) oferecem dois importantes estudos de caso sobre determinações sociolinguísticas em construções alternativas de genitivo no inglês (em *of* e em *'s*) e ditransitivas em neerlandês, respetivamente.

#### 4. O SIGNIFICADO DA VARIAÇÃO LINGUÍSTICA

Num modelo linguístico baseado no uso da língua, os falantes têm uma representação cognitiva das interações comunicativas em que participam. Essa representação cognitiva constitui o seu próprio conhecimento da língua. Faz obviamente parte deste conhecimento a representação cognitiva que os falantes fazem da variação linguística que inevitavelmente caracteriza a comunidade linguística a que pertencem. O significado da variação é pois o sentido que os falantes dão à variação linguística. Os falantes percebem as diferenças letais, categorizam as variedades letais e avaliam atitudinalmente essas variedades. Como qualquer processo de categorização linguística, a categorização da variação letal é naturalmente um processo cognitivo, um processo semântico.

Cabe a um modelo linguístico baseado no uso não só o estudo da variação na produção de uma língua, mas também o estudo da variação na percepção e nas atitudes dos falantes, isto é, o estudo da representação cognitiva da variação letal e dos seus efeitos no uso linguístico e na variação linguística. As questões de investigação incluem saber (i) como é que os falantes percebem as diferenças letais, que modelos cognitivos e culturais utilizam para categorizar a variação letal e como é que avaliam as variedades letais; (ii) até que ponto os fatores perceptivos e atitudinais influenciam a variação e a mudança linguísticas, ou até que ponto as distâncias linguísticas subjetivas (distâncias percebidas, atitudes linguísticas) se correlacionam com as distâncias linguísticas objetivas (observáveis no comportamento linguístico dos falantes); e (iii) até que ponto as distâncias linguísticas objetivas e as atitudes linguísticas influenciam a inteligibilidade mútua de variedades nacionais<sup>4</sup>.

#### 5. A SOCIOLINGUÍSTICA COGNITIVA

A Sociolinguística Cognitiva é uma extensão emergente da Linguística Cognitiva como modelo orientado para o significado e baseado no uso, que pretende investigar a interrelação entre os aspetos sociais e conceptuais da variação letal através de avançados métodos empíricos quantitativos e multivariacionais (Kristiansen & Dirven 2008, Geeraerts, Kristiansen & Peirsman 2010, Silva 2008a, 2009, para uma introdução). Representa a convergência de interesses de investigação da Sociolinguística e da Linguística Cognitiva e contribui quer para integrar na perspetiva e na agenda da Linguística Cognitiva os aspetos sociais da linguagem quer para incorporar na agenda da Sociolinguística os aspetos conceptuais da variação letal.

No contexto do desenvolvimento dos estudos linguísticos das últimas décadas, a Sociolinguística Cognitiva representa o derradeiro esforço de recontextualização da linguagem levada a cabo pela Linguística Cognitiva (não só a inclusão das bases cognitivas e experienciais dos falantes ou contexto cognitivo e do nível interacional do uso linguístico ou contexto situacional, como também a integração do ambiente sociocultural da língua ou contexto social), em resposta à descontextualização (cognitiva, situacional e sociocultural) da gramática, exemplarmente efetuada pela Gramática Gerativa<sup>5</sup>. Deste modo, a Sociolinguística Cognitiva constitui um modelo da referida des-sistematização necessária do sistema linguístico em prol da descrição e explicação de determinada língua como sistema dinâmico complexo.

4. Sobre a inteligibilidade mútua de variedades linguísticas, ver o estudo experimental de Impe, Geeraerts & Speelman (2009).

5. Ver a notável interpretação de Geeraerts (2003, 2010b) do desenvolvimento da linguística moderna como uma sucessão de movimentos de descontextualização (Gramática Gerativa) e recontextualização (inicialmente, Sociolinguística e Pragmática; atualmente e de forma mais integrada e radical, Linguística Cognitiva).

A Sociolinguística Cognitiva traz contributos específicos da maior importância para os dois domínios de investigação da variação letal, referidos nas duas secções anteriores: o domínio da variação do significado e, em particular, a análise dos vários modos de interação entre o significado e as outras fontes de variação linguística e a análise das correlações entre os fatores conceptuais e os fatores sociais da variação; e o domínio do significado da variação ou representação cognitiva da variação letal, nas suas componentes de perceção, categorização e avaliação atitudinal da heterogeneidade interna de uma língua e na sua expressão sob a forma de modelos cognitivos culturais da variação letal. O objeto principal da Sociolinguística Cognitiva é a variação intralinguística ou variação letal em todas as suas formas e dimensões, incluindo também a variação nacional de línguas pluricêntricas (Silva, Torres & Gonçalves 2011), mas ela investiga também a variação interlinguística e os modelos cognitivos e culturais subjacentes a atitudes linguísticas, políticas de língua e ideologias. A Sociolinguística Cognitiva contribui, fundamentalmente, para o modelo multifatorial da gramática, acima referido como necessário para a descrição da língua como sistema dinâmico complexo e para a análise da variação letal inerente a qualquer expressão linguística.

Para o efeito, a Sociolinguística Cognitiva dispõe de modelos descritivos da Linguística Cognitiva. Entre outros, destacam-se a teoria do protótipo (Geeraerts 1985, 1997, Taylor 1995), útil para a descrição de qualquer tipo de variação letal e para a representação cognitiva da variação; os modelos de semântica cognitiva e de gramática cognitiva (Langacker 1987, 1991, 2008; Talmy 2000) e, em particular, a investigação sobre operações de perspetivação conceptual (pontos de vista, construções de ponto de referência, perfilação, subjetificação, etc.), de grande interesse para as questões da variação do significado e da equivalência semântica; as explorações da corporização (“embodiment”) e da situacionalidade sociocultural (“sociocultural situatedness”) da linguagem, para as correlações entre os aspetos conceptuais e sociais da variação letal; a teoria da metáfora conceptual (Lakoff & Johnson 1980, 1999) e a teoria dos modelos cognitivos culturais (Holland & Quinn 1987; Dirven, Frank & Pütz 2003), para o estudo da categorização da variação letal e das atitudes linguísticas.

## **6. MÉTODOS MULTIVARIACIONAIS E SOCIOLETOMETRIA**

Para a descrição adequada de uma língua como um sistema dinâmico complexo, bem como para a descrição adequada da variação letal na sua multivariacionalidade, são necessários métodos empíricos e quantitativos multivariacionais, sejam métodos de observação de corpus, sejam métodos experimentais, seja a combinação de ambos. São igualmente necessárias técnicas letométricas que permitam calcular distâncias linguísticas entre variedades letais e, por exemplo, medir convergência e divergência diacrónica entre variedades letais e estratificação interna sincrónica de variedades letais.

Relativamente à elicitación e experimentação, o método do corpus tem a vantagem de permitir observar o uso real da língua. Com efeito, o que os falantes pensam que fazem com a língua pode não coincidir com o que eles realmente fazem com a língua. Além disso, e argumentando sobre a complementaridade dos métodos empíricos, a análise de corpus fornece uma base empírica consistente para investigações experimentais. Mas uma análise de corpus adequada implica grandes quantidades de dados de corpus, bem como análise estatística, análise quantitativa e multivariacional e técnicas sofisticadas. Não basta uma análise ilustrada por um corpus; é necessária uma análise efetiva de corpus. Por outro lado, a investigação sociolinguística beneficiará sempre com a utilização de técnicas de inquéritos, técnicas de experimentação da psicologia, ou mesmo técnicas da análise neurofisiológica.

No domínio da letometria ou medidas de distâncias linguísticas entre variedades letais, podem distinguir-se três subdomínios: métodos dialetométricos, estilométricos e socioletométricos. A tradição mais clássica da dialetometria tem desenvolvido técnicas avançadas para o cálculo de distâncias linguísticas entre dialetos (Goebl 2006, Nerbonne & Kretzschmar 2003), mais frequentemente para o cálculo de distâncias fonéticas e fonológicas, mas já com algumas aplicações a distâncias lexicais e sintáticas. Uma outra tradição letométrica encontra-se nos estudos sobre os registos linguísticos e consiste no cálculo das diferenças estilísticas entre registos linguísticos. Neste domínio da estilometria, destaca-se a chamada análise multidimensional de Biber (1995). Mais recentemente e, como extensão da investigação letométrica existente, a socioletometria compreende métodos que permitem medir distâncias entre variedades letais e correlacioná-las não só com distâncias no espaço, mas com todos os tipos de variáveis sociolinguísticas. A socioletometria explora, assim, a natureza multifatorial da variação letal, podendo pois combinar as abordagens dialetométrica e estilométrica. Dos poucos grupos de investigação socioletométrica existentes, destaca-se, pelo seu pioneirismo e avançado desenvolvimento, a socioletometria lexical baseada em perfis onomasiológicos (conjuntos de expressões sinónimas alternativas juntamente com as suas frequências no corpus), desenvolvida pela unidade belga “Quantitative Lexicology and Variational Linguistics” para o neerlandês como língua pluricêntrica (Geeraerts, Grondelaers & Speelman 1999; Speelman, Grondelaers & Geeraerts 2003). Medidas de uniformidade baseada em perfis onomasiológicos permitem calcular convergência e divergência entre variedades nacionais. Utilizando também métodos estatísticos multivariacionais aplicados a grandes corpora, o grupo de Gries e Stefanowitsch (Gries & Stefanowitsch 2006, Stefanowitsch & Gries 2006, Stefanowitsch & Gries 2008) tem desenvolvido um modelo colostrucional de grande eficácia na análise das correlações entre a variação lexical e a variação sintática, mas a sua orientação tem sido mais psicolinguística do que sociolinguística.

Para a análise das correlações entre variáveis (fonológicas, lexicais, morfológicas, sintáticas e pragmáticas; sócio-estilísticas e semântico-estruturais; objetivas e subjetivas), podem ser aplicadas várias técnicas quantitativas. Técnicas multivariacionais de “regressão logística” (Pampel 2000) permitem analisar com rigor as correlações entre variáveis e determinar que variáveis são os melhores indicadores de convergência e divergência entre variedades nacionais. Técnicas de visualização, como as “análises de clusters” (Aldenfelder & Blashfield 1984) e as “escalas multidimensionais” (Kruskal & Wish 1978), permitem analisar dissimilaridades entre subcorpora, determinar o poder discriminativo de variáveis e detetar a multidimensionalidade.

No domínio da linguística computacional e das suas aplicações à semântica lexical, são bem conhecidos os “modelos de espaço de palavras” aplicados a grandes corpora, que permitem uma análise automática da distribuição de uma palavra e detetam automaticamente similaridades semânticas entre palavras (Padó & Lapata 2007). Estas técnicas matemáticas podem ser sociolinguisticamente aplicadas na extração automática de sinónimos e na identificação automática de variação lexical entre variedades letais (Heylen, Peirsman, Geeraerts & Speelman 2008; Peirsman, Heylen & Geeraerts 2010; Peirsman, Geeraerts & Speelman 2010). Mais especificamente, a técnica dos “modelos de espaço vetorial” permitirá uma análise mais rigorosa das correlações entre fatores letais e fatores conceptuais de variação semântica no léxico e, por extensão, noutras áreas da língua.

## 7. UM EXEMPLO DE SOCIOLINGUÍSTICA COGNITIVA E SOCIOLETOMETRIA DO PORTUGUÊS: CONVERGÊNCIA E DIVERGÊNCIA ENTRE PE E PB

Desenvolvemos um projeto de investigação centrado na questão diacrónica da convergência ou divergência entre o português europeu (PE) e o português brasileiro (PB) nos últimos 60 anos (Silva 2005b, 2008a, b, 2010, 2011a), cujos métodos e resultados passamos a apresentar muito sumariamente. O projeto utiliza o método onomasiológico de estudo da variação letal, mais precisamente a variação onomasiológica formal entre palavras e construções denotacionalmente sinónimas, referida na secção 3, e métodos socioletométricos baseados em perfis onomasiológicos de conceitos e funções, referidos na secção anterior. A análise baseia-se no corpus CONDIVport, em construção, atualmente com 4 milhões de palavras do registo formal e 15 milhões do registo informal, disponibilizado pela Linguatca (Silva 2008b). O corpus está estruturado na base de três variáveis: (i) geográfica, com textos de Portugal e do Brasil (sobretudo São Paulo e Rio de Janeiro); (ii) diacrónica, das décadas de 50, 70 e 2000; e (iii) estilística, do registo formal de jornais e revistas e do registo informal de chats da internet e de etiquetas de roupas de lojas de vestuário.

A investigação sociolexicológica já realizada (Silva 2010) com base em 43 conceitos nominais – 21 perfis onomasiológicos do campo lexical do futebol e 22 perfis onomasiológicos do campo lexical do vestuário – e em largos milhares de observações do uso destes termos (183 de futebol e 264 de vestuário) no corpus CONDIVport (jornais de desporto e revistas de moda, chats associados a clubes de futebol e etiquetas de roupas de lojas de vestuário) permite concluir que a hipótese da divergência entre PE e PB 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 mais próximos da realidade sociolinguística; a ligeira convergência no campo do futebol será um efeito da globalização e da padronização do vocabulário do futebol.

Os conceitos de futebol estudados são: 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 e TREINADOR. Os conceitos de vestuário são: 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 M/F, CONJUNTO F, FATO M, JAQUETA M/F, JEANS M/F, SAIA F, T-SHIRT M/F e VESTIDO F. A título de exemplo, o perfil onomasiológico de AVANÇADO compreende os termos alternativos atacante, avançado, avante, dianteiro, forward, ponta-de-lança e o perfil onomasiológico de BLUSA F compreende os termos alternativos blouse, blusa, blusinha, bustier, camisa, camisa-body, camisão, camiseiro(inho), camiseta/e, (blusa) chémissier, (blusa) chemisiê.

Como exemplo de aplicação da medida de convergência e divergência entre variedades, a Tabela 1 apresenta as percentagens da medida de uniformidade (U) em relação ao perfil onomasiológico de AVANÇADO nas bases de dados de Portugal (P) e do Brasil (B) entre 1950 e 1970. A medida U obtém-se pela soma das frequências relativas mais pequenas de cada termo alternativo. O aumento de uniformidade entre PE e PB de 16,9% para 28,8% indica convergência relativamente ao perfil AVANÇADO (uma diminuição de uniformidade indicaria divergência).

AVANÇADO	P50	B50	U	P70	B70	U
atacante	8,8	36,6		13,6	73,8	
avanzado	71,6	0,9		47,4	0,0	
avante	0,0	48,9		0,0	11,0	
dianteiro	19,2	6,8		20,1	0,7	
forward	0,1	5,2		0,0	0,0	
ponta de lança	0,3	1,5		19,0	14,5	
			16,9			28,8

Tabela 1. Uniformidade (U) do perfil AVANÇADO entre PE e PB (1950-1970)

A Tabela 2 apresenta as percentagens de uniformidade interna (I), isto é, dentro de uma única variedade linguística, para o mesmo perfil onomasiológico de AVANÇADO nos anos 50: a uniformidade interna é maior no PE (I 55,8%) do que no PB (I 38,1%). Este resultado deve-se aos dois fatores que contribuem para determinar a uniformidade interna: por um lado, P50 apresenta um único termo claramente dominante, ao passo que B50 apresenta dois termos dominantes; por outro lado, há mais termos alternativos frequentes em B50 do que em P50. A medida I obtém-se pela soma das frequências relativas ao quadrado de cada termo do perfil.

AVANÇADO	P50	I	B50	I
atacante	8,8	77,8	36,6	1340,7
avanzado	71,6	5128,8	0,9	0,9
avante	0,0	0,0	48,9	2393,5
dianteiro	19,2	369,2	6,8	45,8
forward	0,1	0,0	5,2	27,4
ponta de lança	0,3	0,1	1,5	2,4
		55,8		38,1

Tabela 2. Uniformidade interna (I) do perfil AVANÇADO em PE e PB em 1950

As Figuras 1 e 2 sistematizam os resultados da análise sociolinguística. São apresentadas as percentagens de uniformidade no corpus de futebol (90.202 observações) e no corpus de vestuário (12.451 observações) nos três períodos estudados (1950-1970-2000). O primeiro número de cada par de resultados representa a uniformidade não-ponderada (U/I) e o segundo a uniformidade ponderada (U'/I'), sendo U'/I' mais significativo do que U/I; os números nas linhas horizontais indicam a percentagem de uniformidade (U/U') entre PE e PB nos anos 50, 70 e 2000; os números nas linhas verticais e diagonais indicam a percentagem de uniformidade (U/U') de período para período e entre diferentes períodos; e os números associados a cada variedade e período indicam a percentagem de uniformidade interna (I/I').

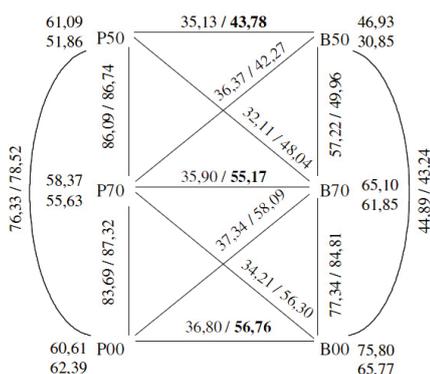


Figura 1. Termos de futebol

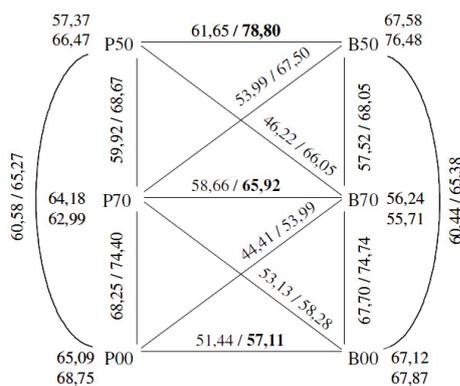


Figura 2. Termos de vestuário

Relativamente aos termos de futebol, as percentagens de  $U'$  mostram convergência de 1950 para 1970 ( $U'$  aumenta 11,39%, de 43,78% em 50 para 55,17% em 70) e estabilidade a partir de então, mas as de  $U$  não indicam qualquer alteração significativa. Ao mesmo tempo, os números sugerem uma grande distância entre as duas variedades nos três períodos. Quanto aos termos de vestuário, as percentagens de  $U$  e  $U'$  evidenciam divergência ao longo do tempo:  $U'$  diminui 12,88% de 1950 para 1970 (de 78,8% para 65,92%) e 8,81% de 1970 para 2000 (de 65,92% para 57,11%).

As linhas verticais e diagonais mostram que não há nenhuma orientação específica de uma variedade em relação à outra: PE e PB divergem um do outro no vocabulário do vestuário; as maiores mudanças do PB no vocabulário de futebol não significam uma aproximação ao PE e a influência do PB no PE no vocabulário do futebol é menor do que o que se esperava.

Um terceiro resultado, expresso nas percentagens de uniformidade interna ( $I/I'$ ), evidencia maiores mudanças no PB em ambos os campos lexicais: no vocabulário de futebol, as percentagens de  $I$  e  $I'$  indicam um grande aumento de uniformidade interna no PB (de  $I$  46,93% /  $I'$  30,85% em 50 para  $I$  65,1% /  $I'$  61,85% em 70 e  $I$  75,8% /  $I'$  65,77% em 2000), ao passo que as alterações no PE são menores; no vocabulário de vestuário, as percentagens de  $I$  e  $I'$  mostram mais alterações no PB (de  $I$  67,58% /  $I'$  76,48% em 50 para  $I$  56,24% /  $I'$  55,71% em 70 e  $I$  67,12% / 67,87% em 2000) do que no PE. Será esta maior mutabilidade do PB o efeito da sua maior complexidade externa, da sua maior variação social ou de um atraso de padronização? Provavelmente um pouco de tudo isto.

Um quarto resultado é apresentado nas Tabelas 3 e 4 e diz respeito à influência dos estrangeirismos nas duas variedades nacionais, nos três períodos estudados, com as percentagens de anglicismos ( $A'$ Ingl), francesismos ( $A'$ Fr) e todos os estrangeirismos ( $A'$ estrang). A medida  $A/A'$  permite calcular a proporção de termos com determinada característica (neste caso, estrangeirismo) no perfil onomasiológico de um conceito e nos perfis onomasiológicos de um conjunto de conceitos na amostra estudada.

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

Tabela 3. Estrangeirismos no vocabulário de futebol

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

Tabela 4. Estrangeirismos no vocabulário de vestuário

No vocabulário de futebol, a influência de anglicismos e outros estrangeirismos é claramente maior no PB do que no PE. No vocabulário de vestuário, verifica-se uma diminuição de francesismos e um claro aumento de anglicismos, por razões socioculturais bem conhecidas. A influência de estrangeirismos continua a ser maior no PB, sendo o PB igualmente permeável à influência de francesismos. Estes resultados confirmam a hipótese de uma maior permeabilidade do PB aos estrangeirismos, quer importando-os diretamente, quer adaptando-os. A Tabela 5 evidencia 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.

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

Tabela 5. Adaptações/decalques de anglicismos no vocabulário de futebol

Como último resultado, o vocabulário do vestuário confirma a hipótese de uma distância sincrónica maior entre estrato padrão e estrato subpadrão no PB do que no PE<sup>6</sup>: a uniformidade entre os dados das revistas de moda de 2000 e os dados das etiquetas de peças de roupa de lojas de vestuário é de 57,94% no PB e de 70,99% no PE. No vocabulário do futebol não há esta assimetria entre os dados dos jornais desportivos de 2000 e os dados dos chats associados a clubes de futebol: a uniformidade é de 80,93% no PE e 78,76% no PB, o que sugere padronização dos termos de futebol na linguagem da internet.

Como extensão atual, pretendemos investigar até que ponto variáveis lexicais e não-lexicais se correlacionam como indicadores de convergência/divergência entre PE e PB. Foram já analisadas 15 variáveis construcionais (Silva 2011a), designadamente construções preposicionais, construções completivas infinitivas e construções adjetivas, assim distribuídas:

- (i) 10 perfis preposicionais (1.730 observações): falar de/sobre/acerca de/em; Nmental acerca de/sobre; interesse em/por; precisar, necessitar de/Ø; ansioso de/para/por; apressar-se a/em/para; convencer, esquecer-se, lembrar-se, recordar-se, admirar-se, duvidar, avisar de que/que; convencido de que/que; gostar, falar, precisar, necessitar relativa/relativa cortadora; esquecer, lembrar de reflexo/não-reflexo;

6. O estrato padrão corresponde aos dados dos jornais desportivos e das revistas de moda portuguesas e brasileiras, ao passo que o estrato subpadrão (não estritamente dialetal, mas de nível intermédio) está aqui representado nos dados dos chats associados a clubes de futebol portuguesas e brasileiras e nos dados das etiquetas de peças de roupa em lojas de vestuário de Portugal e do Brasil.

(ii) 3 construções de complemento infinitivo dos verbos causativos fazer, mandar, deixar e perceptivos ver, ouvir, sentir (1.385 observações): VSV (A Maria fez/deixou/mandou/viu os miúdos correrem), VOV (A Maria fez/deixou/mandou/viu os miúdos correr) e VV (A Maria fez/deixou/mandou/viu correr os miúdos)<sup>7</sup>;

(iii) 2 construções adjetivas com verdadeiro, falso, bonito, lindo, recente (763 observações): N + A (ideia falsa) e A + N (falsa ideia)<sup>8</sup>.

Os resultados mostram que as construções preposicionais (1), infinitivas (2) e adjetivas (3) replicam os resultados dos termos de vestuário, confirmando assim a hipótese da divergência entre PE e PB.

(1)  $U'(P50,B50) 68,74\% > U'(P70,B70) 54,39\% > U'(P00,B00) 48,64\%$

(2)  $U'(P50,B50) 72,62\% > U'(P70,B70) 56,51\% > U'(P00,B00) 50,33\%$

(3)  $U'(P50,B50) 95,83\% > U'(P70,B70) 94,54\% > U'(P00,B00) 83,73\%$

## 8. CONCLUSÃO

Um modelo linguístico baseado no uso da língua implica o abandono da concepção idealista de língua como sistema homogêneo, popularizada por Saussure e por Chomsky, e o reconhecimento da inevitável heterogeneidade de uma língua e da complexa relação dialética entre sistema e uso. Um modelo linguístico baseado no uso implica, fundamentalmente, a visão da língua como sistema dinâmico complexo, a visão sociocognitiva da interação entre os aspetos sociais e os aspetos conceptuais do sistema dinâmico complexo e a compreensão das variedades letais como categorias estruturadas na base de protótipos.

A implicação descritiva de um modelo linguístico baseado no uso é a inclusão da variação letal no estudo de qualquer expressão linguística. Esta inclusão deve dar conta da interação entre os três principais fatores de variação linguística: significado, forma e contexto. A variação lexical e a variação construcional compreendem vários tipos de variação: a variação semasiológica entre referentes e/ou sentidos de determinada palavra ou construção, a variação onomasiológica conceptual entre quase-sinónimos, a variação onomasiológica formal entre sinónimos denotacionais e a variação contextual. Estes diferentes tipos de variação semântica condicionam-se mutuamente. Qualquer fenómeno linguístico é, por natureza, multifatorial, sendo o resultado da correlação de fatores conceptuais, estruturais e sociais. A implicação metodológica, para se poder dar conta da variação letal intrínseca a qualquer expressão linguística e da correlação entre os fatores conceptuais, estruturais e sociais, é a implementação de métodos quantitativos multivariacionais avançados, como os métodos socioletométricos. Como ilustração empírica, vimos que o método onomasiológico do estudo da variação letal (sinónimos denotacionais lexicais e construcionais) e os métodos socioletométricos

7. Ver a análise do significado (valor conceptual) e do uso destas três construções no PE e no PB no estudo cognitivo e empírico de Silva (2005a).

8. Pretendemos ainda analisar outras variáveis não-lexicais, nomeadamente morfológicas e sintáticas. Entre as variantes morfológicas, contam-se nós vs. a gente como pronomes de 1.pl, seu vs. dele como possessivo de 3.sg, variação de concordância nominal e verbal com a gente e alternâncias sufixais, como -ção vs. -mento em substantivos deverbais. As variantes sintáticas incluem a alternância entre infinitivo simples e flexionado (Incentivou os trabalhadores a fazer vs. fazerem greve), a alternância (in)transitiva (O carro furou o pneu vs. O pneu do carro furou), construções intransitivas com/sem se (O departamento reuniu-se vs. reuniu), presença/ausência de preposição em construções relativas preposicionadas e alternância com onde (não há lugar em que/que/onde a polícia não entre), a alternância entre sujeito preenchido e sujeito nulo e a alternância de ordem SV e VS.

baseados em perfis onomasiológicos de medição de convergência e divergência entre variedades letais constituem importantes contributos para o estudo do pluricentrismo da língua portuguesa.

No plano metateórico do desenvolvimento da Linguística, um modelo linguístico baseado no uso da língua implica uma desconstrução radical da noção tradicional de sistema linguístico, uma des-sistematização do sistema linguístico quer externamente (do ponto de vista da variação linguística), conduzindo a uma conceção da língua como sistema dinâmico complexo, quer internamente (do ponto de vista da estrutura gramatical), no sentido da construção de um modelo multifatorial da gramática. A emergente Sociolinguística Cognitiva, extensão inevitável da Linguística Cognitiva como modelo orientado para o significado e para o uso da língua, representa a resposta mais convincente a este desafio de des-sistematização do sistema linguístico.

## **SYSTEM AND VARIATION: HOW SYSTEMATIC CAN THE LINGUISTIC SYSTEM BE IN A USAGE-BASED MODEL?**

### **ABSTRACT**

This paper argues that a usage-based model implies a concept of language as a complex dynamic system which correlates social and cognitive aspects. In descriptive terms, a usage-based model implies the inclusion of language-internal variation in the study of any linguistic expression and its correlation with conceptual meaning. In methodological terms, a usage-based model implies using empirical multivariational methods, such as sociolectometrical methods. On a metatheoretical level, a usage-based model implies the de-systematization of the linguistic system and the construction of a multifactorial model of the grammar. It shows how Cognitive Sociolinguistics, an extension of Cognitive Linguistics, as a model orientated towards meaning and use is able to address these implications. To illustrate this argument, the results of socio-cognitive and sociolectometrical research into divergence between European and Brazilian Portuguese will be presented.

**KEY WORDS:** usage-based model, complex dynamic system, Cognitive Sociolinguistics, language variation, European and Brazilian Portuguese

## REFERÊNCIAS BIBLIOGRÁFICAS

- Aldenfelder, Mark & Roger Blashfield (1984). *Cluster Analysis*. Newbury Park/London/New Delhi: Sage Publications.
- Baldinger, Kurt (1964). Sémasiologie et onomasiologie. *Revue de Linguistique Romane* 28: 249-272.
- Biber, Douglas (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- Chambers, Jack K. & Peter Trudgill (1998). *Dialectology*. Cambridge: Cambridge University Press.
- Colleman, Timothy (2010). *Lectal variation in constructional semantics: "Benefactive" ditransitives in Dutch*. In: Dirk Geeraerts, Gitte Kristiansen & Yves PEIRSMAN (eds.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 191-221.
- Dirven, René, Roslyn Frank & Martin Pütz (eds.) (2003). *Cognitive Models in Language and Thought: Ideology, Metaphors, and Meanings*. Berlin/New York: Mouton de Gruyter.
- Geeraerts, Dirk (1985). *Paradigm and Paradox. Explorations into a Paradigmatic Theory of Meaning and its Epistemological Background*. Leuven: Leuven University Press.
- Geeraerts, Dirk (1997). *Diachronic Prototype Semantics. A Contribution to Historical Lexicology*. Oxford: Clarendon Press.
- Geeraerts, Dirk (2003). Decontextualising and recontextualising tendencies in 20th century linguistics and literary theory. In: Ewald Mengel, Hans-Jörg Schmid & Michael Steppat (eds.), *Anglistentag 2002 Bayreuth*. Trier: Wissenschaftlicher Verlag, 369-379.
- Geeraerts, Dirk (2010a). Schmidt redux: How systematic is the linguistic system if variation is rampant? In: Kasper Boye & Elisabeth Engeberg-Pedersen (eds.), *Language Usage and Language Structure*. Berlin/New York: Mouton de Gruyter, 237-262.
- Geeraerts, Dirk (2010b). Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In: Elzbieta Tabakowska, Michal Choinski & Lukasz Wiraszka (eds.), *Cognitive Linguistics in Action: From Theory to Application and Back*. Berlin/New York: Mouton de Gruyter, 71-102.
- Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema (1994). *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman (1999). *Convergentie en divergentie in de Nederlandse woordenschat*. Amsterdam: Meertens Instituut.
- Geeraerts, Dirk & Hubert Cuyckens (eds.) (2007). *The Oxford Handbook of Cognitive Linguistics*. Oxford/New York: Oxford University Press.

- Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman (eds.) (2010). *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter.
- Geeraerts, Dirk & Dirk Speelman (2010). Heterodox concept features and onomasiological heterogeneity in dialects. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 23-39.
- Goebel, Hans (2006). Recent advances in Salzburg dialectometry. In: John Nerbonne & William Kretzschmar (eds.), *Literary and Linguistic Computing, special issue on Progress in Dialectometry: Toward Explanation*. Oxford: Oxford University Press, 411–435.
- Gries, Stefan Th. & Anatol Stefanowitsch (eds.) (2006). *Corpora in Cognitive Linguistics. Corpus-based Approaches to Syntax and Lexis*. Berlin/New York: Mouton de Gruyter.
- Heylen, Kris, Yves Peirsman, Dirk Geeraerts & Dirk Speelman (2008). Modelling word similarity. An evaluation of automatic synonymy extraction algorithms. In: *Proceedings of the Language Resources and Evaluation Conference (LREC-2008)*. Marrakech, Marrocos.
- Holland, Dorothy & Naomi Quinn (eds.) (1987). *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press.
- Impe, Leen, Dirk Geeraerts & Dirk Speelman (2009). Mutual intelligibility of standard and regional Dutch language varieties. *International Journal of Humanities and Arts Computing* 2: 101-117.
- Kristiansen, Gitte & René Dirven (eds.) (2008). *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter.
- Kruskal, Joseph & Myron Wish (1978). *Multidimensional Sclaing*. Newbury Park/London/New Delhi: Sage Publications.
- Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William (1978). Where does the sociolinguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44. Austin: Southwest Educational Development Laboratory.
- Lakoff, George & Mark Johnson (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.
- Lakoff, George & Mark Johnson (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar. Vol. 2: Descriptive Application*. Stanford: Stanford University Press.

- Langacker, Ronald W. (2008). *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.
- Lavandera, Beatriz (1978). Where does the sociolinguistic variable stop? *Language in Society* 7: 171-183.
- Nerbonne, John & William Kretzschmar (2003). Introducing computational methods in Dialectometry. In: John Nerbonne & William Kretzschmar. (eds), *Computational Methods in Dialectometry, Special issue of Computers and the Humanities*. Oxford: Oxford University Press, 245–255.
- Padó, Sebastian & Mirella Lapata (2007). Dependency-based construction of semantic space models. *Computational Linguistics* 33 (2): 163-199.
- Pampel, Fred (2000). *Logistic Regression. A Primer*. Thousand Oaks: Sage Publications.
- Peirsman, Yves, Dirk Geeraerts & Dirk Speelman (2010). The automatic identification of lexical variation between language varieties. *Natural Language Engineering* 16(4): 469–490.
- Peirsman, Yves, Kris Heylen & Dirk Geeraerts (2010). Applying word space models to sociolinguistics. Religion names before and after 9/11. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 111-137.
- Silva, Augusto Soares da (1999). *A Semântica de DEIXAR*. Uma Contribuição para a Abordagem Cognitiva em Semântica Lexical. Lisboa: Fundação Calouste Gulbenkian e Fundação para a Ciência e a Tecnologia.
- Silva, Augusto Soares da (2005a). Revisitando as construções causativas e perceptivas do Português: significado e uso. In: Inês Duarte & Isabel Leiria (eds.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Associação Portuguesa de Linguística, 855-874.
- Silva, Augusto Soares da (2005b). Para o estudo das relações lexicais entre o Português Europeu e o Português do Brasil: Elementos de sociolexicologia cognitiva e quantitativa do Português. In: Inês Duarte & Isabel Leiria (eds.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Associação Portuguesa de Linguística, 211-226.
- Silva, Augusto Soares da (2006). *O Mundo dos Sentidos em Português: Polissemia, Semântica e Cognição*. Coimbra: Almedina.
- Silva, Augusto Soares da (2008a). Integrando a variação social e métodos quantitativos na investigação sobre linguagem e cognição: para uma sociolinguística cognitiva do português europeu e brasileiro. *Revista de Estudos da Linguagem* 16 (1): 49-81.
- Silva, Augusto Soares da (2008b). O corpus CONDIV e o estudo da convergência e divergência entre variedades do português. In: Luís Costa, Diana Santos e Nuno CARDOSO (eds.), *Perspectivas sobre a Linguateca / Actas do Encontro Linguateca: 10 anos*. Linguateca, 25-28. <http://www.linguateca.pt/LivroL10/Cap04-Costaetal2008-Silva.pdf>

Silva, Augusto Soares da (2009). A Sociolinguística Cognitiva: razões e objecto de uma nova área de investigação linguística. *Revista Portuguesa de Humanidades – Estudos Linguísticos* 13 (1): 191-212.

Silva, Augusto Soares da (2010). Measuring and parameterizing lexical convergence and divergence between European and Brazilian Portuguese. In: Dirk Geeraerts, Gitte KRISTIANSEN & Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 41-83.

Silva, Augusto Soares da (2011a). Para a abordagem socioletométrica do pluricentrismo do português europeu e brasileiro: dos indicadores lexicais aos construcionais e atitudinais. In: Augusto Soares da Silva, Amadeu Torres & Miguel Gonçalves (orgs.), *Línguas Pluricêntricas: Variação Linguística e Dimensões Sociocognitivas / Pluricentric Languages: Linguistic Variation and Sociocognitive Dimensions*. Braga: Aletheia, Publicações da Faculdade de Filosofia da Universidade Católica Portuguesa, 573-592.

Silva, Augusto Soares da (2011b). Palavras, significados e conceitos: o significado lexical na mente, na cultura e na sociedade. *Cadernos de Letras da UFF – Dossiê: Letras e Cognição* 41: 27-53.

Silva, Augusto Soares da, Amadeu Torres & Miguel Gonçalves (orgs.) (2011). Línguas Pluricêntricas: Variação Linguística e Dimensões Sociocognitivas / *Pluricentric Languages: Linguistic Variation and Sociocognitive Dimensions*. Braga: Aletheia, Publicações da Faculdade de Filosofia da Universidade Católica Portuguesa.

Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37: 317-337.

Stefanowitsch, Anatol & Stefan Th. Gries (2006). *Corpus-based Approaches to Metaphor and Metonymy*. Berlin/new York: Mouton de Gruyter.

Stefanowitsch, Anatol & Stefan Th. Gries (2008). Channel and constructional meaning: A collocation case study. In: Gitte Kristiansen & René Dirven (eds.). *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter, 129-152.

Szmrecsanyi, Benedikt (2010). The English genitive alternation in a cognitive sociolinguistics perspective. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter, 141-166.

Talmy, Leonard (2000). *Toward a Cognitive Semantics*. Vol. I: Concept Structuring Systems. Vol. II: Typology and Process in Concept Structuring. Cambridge, Mass.: The MIT Press.

Taylor, John R. (1995). *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Oxford University Press.