

Artigos

O digital na dialetologia do português: apreciação crítica, desafios presentes e apresentação de projeto

The digital in the dialectology of Portuguese: critical appreciation, present challenges, and project introduction

Fernando Brissos*

RESUMO

A transição digital é um fenómeno transversal à sociedade contemporânea e a dialetologia não pode ficar excluída do processo, o que não é o mesmo que dizer que deve abdicar de prioridades bem estabelecidas. No caso do português, o foco deverá estar em (i) processar digitalmente materiais cujo prazo de validade se pode perder e (ii) dar, de forma paralela e consequente, o passo para o paradigma 2.0, i.e. a utilização do canal digital para obter e tratar dados linguísticos primários em larga escala e ligados em rede. Neste trabalho discutimos a utilidade da dialetologia digital para o estudo presente do português e apresentamos um projeto pioneiro, o aplicativo Dialetopédia, que mostra o que se pode fazer e efetivamente se faz hoje, no domínio da dialetologia digital, em língua portuguesa. Concluimos que a conjugação dos planos 1.0 e 2.0 da transição digital é o melhor caminho para a dialetologia do português atual; uma transição eficiente só se fará com a resolução efetiva de vários projetos que correm no paradigma anterior.

Palavras-chave: *dialetologia digital; lusofonia; dialetologia brasileira; dialetologia portuguesa*

*Centro de Linguística e Faculdade de Letras da Universidade de Lisboa. <https://orcid.org/0000-0002-2525-1987>. E-mail: fernandobrissos@edu.ulisboa.pt

ABSTRACT

The digital transition is cutting across contemporary society. Dialectology stands only to benefit from this process, which is not to imply giving up on well-established priorities. In the case of Portuguese, the main concern should be to (i) digitally process resources in danger of becoming outdated, and (ii) take the step towards the 2.0 paradigm, i.e., the use of digital channels to obtain and process primary linguistic data on a large, network-driven scale. In this work we discuss the usefulness of digital dialectology to present-day linguistics of Portuguese and introduce a pioneering project, the Dialetopédia app, which illustrates current possibilities and present work in the domain of Portuguese-language digital dialectology. We conclude that the combination of stages 1.0 and 2.0 of the digital transition is the most effective way to further intersect the dialectology of Portuguese with digitalization. An efficient transition is only achievable through the completion of several projects founded on more traditional approaches.

Keywords: digital dialectology; lusophone world; Brazilian dialectology; Portuguese dialectology

1. Introdução

A dialetologia portuguesa forma um bom prisma para avaliar os prós e os contras do recente domínio de investigação que é a dialetologia digital (i.e. o recurso sistemático a tecnologias digitais para o estudo dos dialetos; densificaremos este conceito adiante). Por um lado, no estudo do português é particularmente evidente que a digitalização da dialetologia depende de um processo mais geral de digitalização da sociedade — ou seja, os estudos dialetológicos aproveitam um recurso que impende sobre eles, não o criam. Por outro lado, a quase nula tradição que as potencialidades mais recentes da dialetologia digital têm no português coloca o investigador perante o dilema, relativamente frequente em história da ciência, de dar o segundo passo sem ter terminado o primeiro — i.e. otimizar a digitalização de algo que ainda não foi sequer digitalizado.

O primeiro aspeto é facilmente entendível: todos ouvimos falar hoje, recorrentemente, da digitalização da sociedade como um desígnio geral para o século XXI, que deve abranger todos os domínios da vida pública (mesmo, parece, muitos domínios da vida privada, como o desenvolvimento dos metaversos sugere). A União Europeia, por exemplo, tem a *transformação digital* da sociedade como um dos seus desígnios para a presente década (veja-se o *Guião para a Década Digital*: União Europeia, 2021), e são cada

vez mais os países com *planos* ou *estratégias nacionais de digitalização*. A dialetologia é, então, apenas mais uma área da sociedade que procura aproveitar as enormes potencialidades que a transição digital oferece, e o facto de não estar no lugar do condutor tem implicações nesse aproveitamento. Para compreender melhor essas implicações, porém, necessitamos de discutir o segundo aspeto referido acima.

Podemos considerar que a dialetologia digital tem dois planos fundamentais: a transposição para o domínio informático de *corpora* linguísticos obtidos por outros meios, por um lado, e a utilização de ferramentas digitais para recolher e sistematizar dados linguísticos em larga escala (em rede), por outro. Contam-se no primeiro caso a publicação de atlas linguísticos *online* (como p. ex. o *Atlas Linguístico-Etnográfico dos Açores*: Saramago, 2012) ou a constituição de bases de dados informatizadas a partir de informação bibliográfica previamente publicada (e.g. *Tesouro do léxico patrimonial galego e português*, que compila léxico dialetal a partir de todo o tipo de obras que se ocupem do assunto, sejam estudos lexicográficos ou simples monografias de licenciatura: Álvarez, 2009-). No segundo caso ainda quase não existem exemplos em português; analisaremos na secção 2 um projeto em desenvolvimento pelas universidades de Brasília e Lisboa que procura contribuir para mudar esse paradigma, que noutras línguas, como o castelhano, se tem vindo a transformar há já algum tempo (é exemplo clássico o recurso a materiais do Twitter: Estrada & Benito, 2016).

O segundo plano referido é, claro, o mais recente. Ao contrário da mera transposição de informação linguística para suporte eletrónico, que pode provir de *corpora* recolhidos há décadas por meio das metodologias tradicionais (o citado atlas dos Açores é disso um bom exemplo), no segundo plano a própria recolha dos dados é feita em ambiente digital, seja por via direta (como inquéritos linguísticos *online*) ou por via indireta (como os dados obtidos em redes sociais). É, por isso, um passo em frente na digitalização da dialetologia. Mas é proveitoso, para a dialetologia do português, investir fortemente nesse domínio agora? Ou existem coisas por arrumar no curto/médio prazo que possibilitarão um melhor aproveitamento da dialetologia digital 2.0 (como, por exemplo, a conclusão dos projetos dos atlas linguísticos nacionais de Portugal e do Brasil, de que falaremos com detalhe adiante)? Uma vez que a digitalização é um processo transversal e inevitável na sociedade do séc. XXI e que, por isso, a dialetologia não está no lugar do condutor de um veículo que vai a alta velocidade, esta não deverá arriscar-se a tentar ser a barreira intransponível contra uma força

imparável. Deve, no entanto, manter uma perspetiva mais utilitária do que exploratória neste momento. Explicaremos porquê.

2. Análise panorâmica: correlação entre necessidades e virtualidades da dialetologia digital na geolinguística do português

A progressiva aplicação de tecnologias digitais à dialetologia é, como vimos, um processo que transcende este domínio de estudos. Ora, com as imensas possibilidades que esse processo oferece, não deixa também de apresentar alguns riscos que é necessário ter em conta. O maior deles, no caso do português, é incorrer-se, mais uma vez, no esquecimento de um desígnio fundamental: a publicação dos atlas linguísticos de Portugal e do Brasil, lacuna muito substancial da nossa geografia linguística.¹ Depois do atraso significativo na concretização de atlas linguísticos nacionais em ambos os países (no Brasil os trabalhos começaram nos anos 1990 e em Portugal nos anos 1970), agora que as recolhas estão concluídas e os resultados finais quase prontos a serem publicados, importará não deixar mais uma inflexão metodológica atrasar os projetos. Existem precedentes importantes — a divulgação da linguística teórica e da sociolinguística a partir dos anos 1960, que retiraram muita da massa crítica da dialetologia tradicional —, pelo que é decisivo não perder o foco. Com efeito, se a cartografia linguística é um recurso fundamental para conhecer o tecido social de uma dada área ou nação, é-o especialmente no caso do português. As visões de conjunto dos sistemas dialetais de Portugal e do Brasil que são hoje utilizadas de forma quase unânime têm já décadas e dependem de organizações territoriais em larga parte ultrapassadas; falamos, respetivamente, de Cintra (1971, mapa 1²) e Nascentes (1953, mapa 2). Em termos simples, o Brasil de Antenor Nascentes e o Portugal de Lindley Cintra já não existem. Tome-se como exemplo o caso de Portugal, que tem a visão de conjunto mais recente.

1. Veja-se uma breve descrição do atlas linguístico de Portugal, o ALEPG – *Atlas Linguístico-Etnográfico de Portugal e da Galiza*, e do seu contexto em Brissos (2022); uma descrição pormenorizada está disponível em Saramago (2006). O ALiB – *Atlas Linguístico do Brasil* pode ser conhecido facilmente através de conteúdos de vária natureza (incluindo vídeos) disponíveis na sua muito completa página *web*: <https://alib.ufba.br/>. Os dados de ambos os projetos estão, na sua larga maioria, inéditos. Do ALEPG foram publicados os dados concernentes aos Açores (no já citado ALEAç, que é um recorte do ALEPG) e uma pequena parte dos concernentes à Madeira (Nunes *et al.*, 2018), permanecendo a informação relativa ao continente totalmente inédita; do ALiB, foram publicados alguns dados relativos às capitais estaduais (Cardoso *et al.*, 2014).

2. Todos os mapas utilizados no presente trabalho são colocados em anexo no final do texto.

Os dados em que Cintra apoia a sua classificação dialetal foram obtidos nos anos 1950, nas campanhas de inquérito para o ALPI – *Atlas Lingüístico de la Península Ibérica*; o então jovem Cintra foi inquiridor na maior parte dos inquéritos portugueses do projeto.³ Mas Portugal sofreu intensas mudanças no seu tecido demográfico e social desde então: industrializou-se, democratizou-se (terminou uma ditadura de quase meio século em 25/04/1974), perdeu o império (descolonizou as suas possessões em África em 1975, após uma guerra de 14 anos), integrou-se na comunidade de estados europeus (entrou em 01/01/1986 na hoje União Europeia, então Comunidade de Estados Europeus) e, particularmente relevante para a sua estrutura dialetal, litoralizou-se e alfabetizou-se. Veja-se o mapa 3 (com o quadro anexo), que deixa bem visível o intenso processo de litoralização (comumente chamado também de desertificação do interior) que o tecido demográfico de Portugal tem sofrido recentemente: por um lado, foi só no interior que se deu decréscimo populacional contínuo desde o pós-Segunda Grande Guerra; por outro lado, de todas as regiões que tocam na fronteira com Espanha e não tocam no litoral (o interior mais profundo), só a região de Évora (uma das maiores cidades do interior português) não sofreu o referido decréscimo populacional. Se, por outro lado, nos focarmos na alfabetização/escolarização da população, a mudança também tem sido profunda; cf. a Tabela 1, onde está patente a verdadeira inversão que, em apenas meio século, se deu na relação entre ausência e posse de nível de ensino na população residente no país.

Tabela 1 — População residente em Portugal com 15 e mais anos, segundo os Censos, sem nível de ensino, em % da população residente.

Anos	Proporção em %
1960	65,6
1970	35,2
1981	21,9
1991	13,0
2001	9,2

Fonte: PORDATA

3. Apesar de as recolhas de dados terem terminado já nos anos 1950 (tendo-se iniciado nos anos 1930), o atlas linguístico ibérico ainda está maioritariamente inédito. Apenas um volume foi publicado até hoje, com 75 cartas fonéticas (1962), mas desenvolve-se atualmente um projeto no Consejo Superior de Investigaciones Científicas (CSIC) (Madrid) que visa editar *online* todo o material do ALPI (García Mouton *et al.*, 2009-). Uma apresentação breve do ALPI no contexto português pode ser vista em Brissos (2022).

É claro que a investigação geolinguística não pode deixar de acompanhar a evolução da própria língua, quer dizer, a visão de sistema dos dialetos portugueses não pode continuar a depender de dados prévios a evoluções estruturantes da sociedade e, por isso, da língua real. Por isso Castro (2016) refere que a divisão Norte/Sul postulada por Cintra para os dialetos portugueses continentais já não se verifica, mas sim uma divisão litoral/interior;⁴ e por isso Brissos (2018) reviu a classificação de Cintra no noroeste do país, a área-berço da língua, a partir precisamente dos dados do ALEPG (a única proposta de revisão dessa classificação efetuada até hoje a partir de dados linguísticos concretos). Mas falta pôr em cima da mesa todos esses dados, que deverão provocar um efeito dominó no estudo da paisagem geolinguística do país; um atlas é sempre muito mais do que a mera publicação de mapas sobre um determinado assunto, especialmente quando, como o ALEPG, é produto único (não existe nem existiu outro projeto de atlas linguístico nacional) e ferramenta indispensável para resolver importantes questões que pendem há já demasiado tempo.

Esse efeito dominó também é previsível no Brasil, que sofreu não menos intensos movimentos de população desde o pós-guerra. A colonização de áreas historicamente muito pouco povoadas — desde logo, mas não só, o «território incaracterístico» da classificação de Antenor Nascentes (1953) —, a deslocação da capital do país do litoral para o planalto interior (com a fundação de Brasília em 1960), a transição democrática a partir do final dos anos 1980 e as condições inerentes a uma economia de *país em desenvolvimento* fazem do Brasil, reconhecidamente, um país extremamente dinâmico — com toda a instabilidade que isso traz.

É absolutamente fundamental, então, atualizar a nossa visão dos sistemas dialetais de Portugal e do Brasil. E a concretização desse desígnio passa invariavelmente pela dialetologia digital 1.0, i.e. a otimização de *corpora* obtidos fora da *web*. O ALEPG é, mais uma vez, um exemplo pa-

4. Nas próprias palavras do autor, que fala em entrevista: «Em Portugal, há 50 anos, dizia-se que havia dialetos no Norte e no Sul do país que se distinguiam, e no Norte havia mais dialetos que se distinguiam uns dos outros. [...] Havia um panorama que foi cartografado, que é o mapa de dialetos de Lindley Cintra. Um mapa muito semelhante ao das regiões portuguesas de Orlando Ribeiro. São produtos da mesma época e que davam conta da situação nessa altura. Hoje em dia não há isso. Não há a mesma divisão entre Norte e Sul. Há, sim, divisão entre litoral e interior. E o litoral é constituído por uma espécie de megacidade onde vive quase toda a população portuguesa, que vem desde Viana do Castelo até ao Algarve com uma interrupção entre Sines e Lagos. Nesta megacidade predomina um superdialeto que é o português standard – que já não é só o português de Lisboa. É o português da comunicação social, é o português das escolas.»

radigmático. Tendo as recolhas sido efetuadas entre 1973 e 2004, sempre em gravações de sinal analógico, a equipa do atlas desenvolveu nas últimas décadas os trabalhos decorrentes à transição digital do *corpus*: digitalizou todas as gravações, que se encontram disponíveis no Centro de Linguística da Universidade de Lisboa (CLUL), e tem vindo a inserir todas as transcrições fonéticas e ortográficas das perguntas do questionário numa base de dados relacional disponível em *intranet* nas instalações do Centro. A base de dados, para além de possibilitar a extração da totalidade dos dados em ficheiros Word, Excel ou PDF, permite também efetuar pesquisas por itens específicos: campo semântico do questionário, pergunta do questionário, resposta em forma ortográfica ou fonética, inquéritos ou conjuntos de inquéritos, informantes, etc. À data da redação deste artigo, todos os dados relativos ao território português estavam já inseridos na base de dados, faltando apenas os materiais recolhidos em regiões espanholas confinantes com Portugal continental.

O ALiB, projeto mais recente, segue um caminho necessariamente paralelo ao do ALEPG, mas nem todos os *corpora* têm tido a mesma sorte. Um caso particularmente notório é o ILB – *Inquérito Linguístico Boléo*, projeto de geografia linguística desenvolvido na Faculdade de Letras da Universidade de Coimbra (FLUC) por Manuel de Paiva Boléo, ao longo de mais de 30 anos, entre o princípio dos anos 1940 e os anos 1970. Apesar de se tratar do primeiro projeto de geolinguística transversal ao país (i.e. de tipo atlas) feito por uma instituição portuguesa, os seus dados permanecem globalmente inéditos e em formato apenas físico, embora disponíveis para consulta na FLUC. O ILB é, para além disso, um *corpus* massivo, com mais de 2.500 povoações abrangidas e um questionário de mais de 500 perguntas (que, porém, está longe de ter sido sempre aplicado na sua totalidade); na primeira fase do projeto, em 1942, os inquéritos foram efetuados por correspondência, e na segunda fase, entre 1943 e 1974, foram obtidos dados *in loco* por inquiridores de diferente perfil, mas normalmente estudantes universitários.⁵ Se recuarmos no tempo e considerarmos o enorme volume de estudos do pioneiro da dialetologia científica portuguesa, Leite de Vasconcelos (n. 1858, m. 1941), que compreendem mais de meio século de recolhas *in loco* de dados com um perfil de informante extremamente conservador (praticamente inacessível hoje em dia), temos em consideração mais um enorme *corpus* que permanece por aproveitar à luz das potencialidades digi-

5. Pode ver-se uma breve síntese do ILB em Brissos (2012, p. 47-48); descrições detalhadas estão disponíveis em Boléo (1974a; 1974b; 1974c; 1978).

tais. O grosso dos materiais dialetológicos de Leite que foram já publicados consta da sua coletânea *Opúsculos*, nomeadamente: vol. II (Vasconcelos, 1928), relativo ao «dialecto interamnense», i.e. às variedades linguísticas da antiga província do Entre-Douro-e-Minho; vol. IV (Vasconcelos, 1929), relativo às variedades linguísticas periféricas e/ou fronteiriças do português europeu e à classificação dialetal do português europeu continental; e vol. VI (Vasconcelos, 1985; edição póstuma), relativo aos dialetos de Trás-os-Montes e Alto Douro e das Beiras. Permanecem assim por publicar os materiais concernentes ao sul do país e às ilhas e ainda outro material dialetológico que o colecionador compulsivo que era Leite de Vasconcelos nos legou. Ora, se juntarmos os *corpora* de Leite de Vasconcelos (o maior conjunto homogêneo de dados dialetais já publicados sobre o português), do ILB (o projeto com a mais densa rede de inquéritos) e do ALEPG (o atlas linguístico nacional), temos um volume assinalável de dados enquadrado num contínuo histórico que ainda não estudamos porque não podemos. Mas devíamos poder.

A dialetologia do português permite pois evidenciar que não é só no desenvolvimento de ferramentas novas que o paradigma digital pode ser útil. A mera disponibilização de informação geolinguística em formato *online*/informático é também dialetologia digital (em versão 1.0, como dizíamos) — e é também absolutamente crucial, e por isso não pode ser esquecida. A dialetologia digital pode, todavia, ser muito mais do que isso, e como tal deve ser aproveitada. Veremos seguidamente o principal projeto que corre nesse âmbito em português.

3. Análise de caso: o projeto Dialetopédia

O vocabulário do português tem sido tratado essencialmente em obras lexicográficas apontadas a um padrão nacional. Quer isso dizer que não tem havido um tratamento sistemático da variação dialetal da língua nos dicionários e obras afins (e.g. tesouros, vocabulários), por um lado, nem têm esses estudos adotado uma visão integrada das diferentes variedades nacionais da língua, por outro. Mas a língua existe, e vamos dispor de cada vez mais recursos científicos e técnicos (vimos exemplos de uns e outros nas secções anteriores) para fazer face àqueles dois objetivos. Por isso uma equipa composta por investigadores das universidades de Brasília (UnB) e Lisboa (UL) começou há três anos a desenvolver o projeto Dialetopédia, que, em termos simples, pretende construir uma *app* com o essencial da

estrutura lexical das diferentes variedades do português, numa configuração de tipo dicionário enciclopédico. O objetivo último é, portanto, disponibilizar ao grande público um acervo alargado dos lexemas e seus significados na língua portuguesa, com as variações inerentes. O projeto está sediado na UnB e é coordenado, em primeira instância, por Yuko Takano e Abdelhak Razky (UnB) e, em segunda instância, por Fernando Brissos (UL). Outras instituições, porém, têm vindo e vão continuar a juntar-se.

O Dialetopédia ocupa-se então da *lusofonia* como um todo, embora estejamos longe ainda de dispor de recursos dialetais para estudar o português africano ou asiático, por exemplo. O foco a curto e médio prazo são, por isso, as variedades portuguesa e brasileira da língua. E, por estrutura lexical, entender-se-ão palavras simples e complexas e estruturas frásicas análogas; portanto o projeto bebe da lexicologia e da fraseologia. No plano metodológico, enquadra-se não apenas no paradigma tradicional da dialetologia e da geografia linguística (ou geolinguística) mas também no aporte que a sociolinguística tem trazido aos estudos da variação. Fá-lo por duas razões: uma geral, que consiste na necessária atualização de uma disciplina já com mais de século e meio (a dialetologia faz parte das disciplinas pioneiras da linguística); e uma particular, nomeadamente a adaptação que é forçoso fazer da dialetologia/geolinguística tradicional, ciência de base europeia, a variedades coloniais e pós-coloniais como o português brasileiro, africano, etc. É hoje bem sabido que, se um estudo meramente horizontal, apoiado no informante NORM (*non-mobile, older, rural male*), fornece resultados substantivos no estudo de línguas consolidadas e analisadas a partir do seu berço espontâneo (e.g. o português de Portugal ou o castelhano de Espanha), o mesmo não se passa além-mar, onde a variação vertical (entre classes ou grupos sociais) é muito mais premente (e uma simples consequência dos distantes estatutos socioeconómicos dos grupos de indivíduos em causa).

Os materiais linguísticos utilizados pelo projeto são de duas ordens: secundária e primária. Aqueles consistem nos amplos recursos que a tradição geolinguística portuguesa e brasileira tem vindo a criar ao longo das décadas e que, como já vimos, têm também, pontualmente, vindo a ser digitalizados; falamos de atlas linguísticos mas também de *corpora* como o ILB ou monografias dialetais, em qualquer caso referentes apenas ao português contemporâneo (o único período da história da língua que convive com a dialetologia científica). Os materiais primários, por seu lado, consistem em recolhas feitas no âmbito do projeto, que hoje em dia, apesar das novidades introduzidas pela pandemia de Covid-19, se tornam mais

fáceis com as muitas possibilidades digitais de que dispomos. A conjugação de duas ordens de materiais que se reportam ao mesmo período da história da língua justifica-se, desde logo, porque os dados secundários permitirão cobrir simultaneamente (i) uma rede mais densa de locais do que os dados primários passíveis de ser obtidos a curto-médio prazo e (ii) uma gama mais variada de perfis de informante (p. ex. falantes com pouco contacto com o *standard* linguístico, que, como vimos na secção 2, são cada vez mais uma raridade). Em sentido geral, podemos dizer que os dados secundários nos fornecem muita informação; o historiador da língua/dialetólogo não está habituado a rejeitar informação, nem pode facilmente rejeitar materiais.

As recolhas de dados primários serão efetuadas por meio de inquéritos presenciais e virtuais apoiados num questionário semântico-lexical com componente dirigida e semidirigida, cuja estruturação obedece às seguintes fases: *a)* levantamento de lexias simples e complexas a partir das fontes bibliográficas; *b)* elaboração de um questionário preliminar; *c)* aplicação do questionário preliminar a uma amostra das regiões investigadas, o que permite o teste dessa versão do questionário e a sua ampliação com itens que não estão registados na bibliografia; *d)* organização de um questionário piloto, a ser aplicado a uma amostra alargada das regiões consideradas; *e)* configuração do questionário definitivo de cada campo semântico. Os campos semânticos em consideração incluem temas tão variados como culinária, turismo, convívio social, corpo humano, vestuário, utensílios, habitação, crenças e superstições, religiões, costumes, fauna e flora, etc.

Quer as recolhas presenciais quer as virtuais serão dirigidas por um ou mais investigadores nativos da região investigada e com *expertise* no/s campo semântico investigado. O questionário é aplicado a pelo menos oito falantes nativos de cada região, não se procedendo necessariamente, por razões práticas, a uma divisão em células iguais na estratificação social. Essa estratificação inclui: uma faixa etária global que varia de 25 a 80 anos e é subdividida em função da realidade do trabalho de campo; quatro informantes por faixa etária; homens e mulheres; índice de escolaridade adaptado à realidade local e geracional. A estipulação de um perfil de informante tão plural permite não só uma melhor descrição das realidades linguísticas em causa, que são bem diferentes, mas também o aproveitamento dos dados para ulterior análise de aspetos diageracionais, diagenéricos e diastráticos.

A informação linguística recolhida é sistematizada numa base de dados de formato MySQL. O registo de cada lexema é constituído por uma en-

trada que direciona para outras células, que por sua vez são destinadas aos diferentes elementos constitutivos da microestrutura que irá caracterizar o dicionário eletrónico em que a *app* se baseia. A base de dados é atualizada permanentemente pelo segmento da equipa de trabalho competente neste domínio técnico, em articulação com os investigadores que fizeram o trabalho de campo e sob a supervisão dos diretores do projeto.

Uma vez que o projeto se destina a construir uma *app*, o público será o mais abrangente possível, o que tem consequências importantes ao nível da macro e microestrutura do *software*. Deve considerar-se, desde logo, a necessidade de disponibilizar informações relativas à organização da entrada do lexema, estrutura gramatical, variação dialetal, definição e abonação, mas também remissões, ilustrações (imagens, vídeos, caso seja relevante), informações culturais em sentido lato e fonte. Assim, a macroestrutura fundamentar-se-á numa abordagem semasiológica (i.e. da palavra para o conceito) aos lexemas, com configurações onomasiológicas (do conceito para a palavra) superiores por campo semântico. A microestrutura, por sua vez, organiza-se nos seguintes itens específicos: *a)* entrada: a unidade lexical simples ou complexa; *b)* categoria gramatical: apresentação das classes gramaticais das unidades lexicais; *c)* variação dialetal: registo das variações diatópicas de itens lexicais de português falado no mundo; *d)* definição: delimitação do conteúdo semântico da entrada; *e)* informação cultural: esclarecimentos sobre aspetos culturais relativos ao espaço dialetal onde a variante ocorre; *f)* abonação: exemplos retirados de fontes impressas, manuscritas ou orais; *g)* organização de conteúdo multimédia: suportes, sob a forma de imagem, áudio e vídeo, que ilustram a dimensão diatópica da entrada; *h)* remissões: *links* hipertextuais que remetem para outras entradas relacionadas com a entrada principal; *i)* fonte: indicação das referências que sustentam a exposição, sejam retiradas de bibliografia ou das entrevistas efetuadas.

Fazendo uso das potencialidades oferecidas pelo tipo de *software* utilizado, o aplicativo incluirá mapas linguísticos virtuais, criados por meio da ferramenta [SGVCLin] – *Software para geração e visualização de cartas linguísticas* (Romano, Seabra & Oliveira, 2014). Esses mapas podem ser gerados e distribuídos por informação diatópica, diatópica/diagenérica e diatópica/diageracional, permitindo por isso satisfazer necessidades de vários tipos de público (desde o especialista ao utilizador *naïf*).

Um projeto da natureza do Dialetopédia divide-se, naturalmente, em fases, o que, em termos práticos, implica que os campos semânticos não

são todos considerados de uma só vez. A primeira fase, ou versão 01, do projeto ocupa-se do campo semântico relativo à culinária, e cada versão se ocupará de apenas um campo semântico adicional. A versão 01 servirá porém de protótipo e teste genérico à metodologia utilizada, tanto no que toca ao plano linguístico como ao plano informático; daí, precisamente, a escolha por um campo semântico simultaneamente apelativo para o utilizador e desafiante para o investigador, pela elevada quantidade de material que é necessário tratar.

A versão 01 é efetuada de 2021 a, presumivelmente, 2026 e organiza uma base de dados do léxico da culinária no espaço dialetal de Brasil e Portugal. O campo semântico considerado é segmentado em vários subcampos, entre os quais *a)* receitas; *b)* pratos típicos nacionais e regionais; *c)* hábitos alimentares; *d)* higiene alimentar; *e)* utensílios; e outros aspetos a definir eventualmente após o esgotamento da pesquisa bibliográfica e a recolha de dados primários.

O fluxo de trabalho estipulado para a versão 01 está organizado nos 17 momentos indicados a seguir. Uma vez que a versão 01 é prototípica, este fluxo servirá naturalmente de base de trabalho para os fluxos das versões ulteriores da *app*, que neste momento não é produtivo definir.

1. *Seleção das fontes bibliográficas* sobre culinária que subsidiarão o levantamento do *corpus* (dicionários gerais e especializados, documentos enciclopédicos, livros especializados, atlas linguísticos, monografias académicas, *sites* especializados e outros recursos multimédia disponíveis);
2. *Extração das unidades lexicais* simples e complexas dos dicionários do domínio português;
3. *Elaboração e aplicação do questionário* de acordo com o já enunciado;
4. *Seleção das unidades lexicais para a base de dados*, i.e. recolha dos lexemas que servirão de *input* para a constituição do dicionário eletrónico; esta fase conta com ferramentas computacionais de tratamento de *corpora*;
5. *Análise linguística dos dados selecionados*, que envolve a componente dialetológica das unidades lexicais simples e complexas, a componente lexicográfica para a organização dos verbetes e uma análise fraseodialetoal do *corpus*;

6. *Elaboração de critérios linguísticos para a organização da base de dados*: define-se neste momento o conteúdo linguístico de cada célula da base de dados, de modo a facilitar a organização e extração automática de dados;
7. *Elaboração da macro e microestrutura* de acordo com os objetivos gerais da base de dados a ser transposta para o aplicativo (diferente da base de dados prévia em que a maioria dos membros da equipa pode intervir);
8. *Configuração da base-padrão de verbetes*: elabora-se um modelo de verbete completo, que serve de referência para a configuração de todos os outros verbetes do dicionário;
9. *Elaboração dos verbetes*: na posse de um modelo de verbete e de uma base de dados completa, produzem-se os verbetes relevantes para a versão do aplicativo;
10. *Elaboração e inserção do conteúdo multimédia*: são inseridas no aplicativo imagens estáticas ou animadas, áudios e vídeos que ilustrem os conteúdos e, com isso, facilitem o acesso à competência dialetal do falante nativo de ambas as variedades em estudo;
11. *Seleção de ferramentas computacionais*: em parceria com uma equipa técnica da UnB, procede-se à escolha da plataforma computacional adequada para a elaboração da *app*;
12. *Desenho da interface do utilizador*: mais uma vez em parceria com uma equipa técnica, elabora-se uma interface; tem-se como prioridade máxima facilitar a experiência do utilizador;
13. *Alimentação da base de dados*: a equipa linguística passa a alimentar a base de dados do aplicativo;
14. *Desenvolvimento e testes do aplicativo*: a equipa técnica tem agora a responsabilidade de programar e testar os módulos necessários para a *app*, cabendo à equipa linguística o controlo de qualidade dos dados;
15. *Validação dos verbetes*: recorre-se aos investigadores especializados no campo semântico em Portugal e no Brasil para a revisão dos verbetes;
16. *Elaboração de cartas diatópicas*, i.e. mapeamento das unidades monoléxicas e poliléxicas registadas, com as respetivas variantes;
17. *Distribuição da versão 01 do aplicativo*.

A informação compilada ao longo das tarefas linguísticas converge no verbete, que se constitui como um elemento basilar do fluxo de trabalho. O verbete da versão 01 decompõe-se em 13 elementos sequenciais que proporcionam uma cobertura teoricamente completa do objeto de estudo: (i) a entrada, que se refere a lexemas simples ou complexos do campo semântico considerado (culinária), definidos e mapeados na *app*; (ii) a categoria gramatical; (iii) a pronúncia, por meio de ficheiro áudio; (iv) as variantes diatópica, diastrática e diageracional; (v) os usos (técnico, vulgar ou comum, pejorativo, etc.); (vi) definição; (vii) fraseologias relacionadas; (viii) abonação (exemplo e fonte do exemplo); (ix) etimologia (informação genética e histórica do lexema); (x) receitas/modo de preparação, com indicação de especificidades regionais (por exemplo, a variação na preparação e consumo do açaí entre o norte e o sul do Brasil) e informações sobre os tipos de ingredientes utilizados (cujos nome e forma podem variar significativamente entre regiões); (xi) conteúdo multimédia ilustrativo; (xii) informações enciclopédicas e culturais; (xiii) cartas linguísticas, i.e. mapeamento dos lexemas.

Os resultados esperados da versão 01 do Dialetopédia enquadram-se em dois planos: *a*) a disponibilização de uma base de dados sobre as unidades lexicais concernentes ao campo semântico alargado da culinária em português brasileiro e europeu, e *b*) a disponibilização do aplicativo proposto para dispositivos móveis IOS e Android. O aplicativo será, como se referiu, *user-friendly* e destinado ao leque mais alargado possível de público falante de português ou apenas interessado no tema; a interface será totalmente acessível para o utilizador não especializado, os conteúdos apresentados de forma dinâmica e intuitiva e a informação escalonada de acordo com as preferências do utilizador. O carácter abrangente do público-alvo da *app* justifica-se, primariamente, a partir do facto de que a língua é de todos, todos a falam e todos a fazem evoluir; e, secundariamente, pela natureza da própria transição digital (tomada em sentido lato), que se concentra precisamente em tornar cada vez mais recursos mais acessíveis.

4. Conclusão

Tratámos de dois pontos principais neste artigo. Em primeiro lugar, a transição digital da dialetologia do português como caso exemplar de um dilema que é recorrente em história da ciência: adotar uma nova metodologia rapidamente, correndo o risco de perder as prioridades certas, ou adotar uma

postura conservadora, correndo o risco de estagnar. Vimos que a dialetologia do português tira mais proveito, neste momento, da dialetologia digital 1.0 (digitalização de *corpora* já existente e recolhido por meios tradicionais, como os atlas linguísticos); mas também vimos que as potencialidades da dialetologia digital são imensas e é pouco sensato rejeitá-las (e com isso tentar ser, como dizíamos, a barreira intransponível que enfrenta a força imparável). A solução é, portanto, um equilíbrio pragmático: digitalizar *corpora* tradicionais, que existem e não podem ser esquecidos, e, de forma paralela, começar estrategicamente a tirar o proveito devido das potencialidades da dialetologia digital 2.0 (utilização do canal digital para obter dados primários). Os equilíbrios pragmáticos não são tão frequentes assim.

Em segundo lugar, mostrámos um exemplo de incorporação do paradigma 2.0 na dialetologia do português: o projeto Dialetopédia, que tem como objetivo construir uma *app* que torne disponível ao grande público, numa perspetiva comparativa, o essencial do vocabulário das variedades da lusofonia (i.e. do português no mundo). Mais do que mero exemplo, de resto, a Dialetopédia é um projeto verdadeiramente pioneiro na transição para o digital 2.0 em português; por isso dissecámos a estrutura do projeto, tentando mostrar o que, em termos práticos, é possível fazer em dialetologia digital atualmente em português, seja nas suas limitações, como a falta de dados para as variedades do português africano, seja nas suas vantagens, como a plena intercomunicação entre os falantes das várias proveniências. Na experiência da Dialetopédia podemos também ver, noutro sentido, como os resultados da referida transição não vão ser imediatos, o que só nos lembra da necessidade de não descurar os projetos que a geolinguística tradicional nos legou e, nos casos mais resistentes, continuam em curso.

Conseguir ser-se consequente é, portanto, o desafio principal da dialetologia digital do português no início do século XXI: a digitalização é um mundo, mas um mundo que é preciso incorporar na galáxia da geolinguística, e não o contrário.

Agradecimentos

O presente trabalho foi desenvolvido no âmbito das atividades do projeto UIDB/00214/2020, que decorre no Centro de Linguística da Universidade de Lisboa e é financiado pela Fundação para a Ciência e a Tecnologia (FCT) (Governo de Portugal).

Conflito de interesses

Declaramos não ter qualquer conflito de interesse, em potencial, neste estudo.

Disponibilidade dos dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível mediante solicitação ao autor. O conjunto de dados não está publicamente disponível devido a se encontrar em processamento no aplicativo Dialetopédia.

Referências

- AA. VV. (1962). *Atlas Lingüístico de la Península Ibérica*, vol.1, Fonética. CSIC.
- Álvarez, R. (Coord.) (2009-). *Tesouro do Léxico Patrimonial Galego e Português*. Instituto da Língua Galega. <http://ilg.usc.es/Tesouro/pt/> (consulta: 16/01/2023).
- Boléo, M. de Paiva (1974a). O estudo dos dialectos e falares portugueses: um inquérito linguístico. In M. de Paiva Boléo, *Estudos de Linguística Portuguesa e Românica*, vol.1, Dialectologia e História da Língua, t.1 (pp. 1–43). Universidade de Coimbra.
- Boléo, M. de Paiva (1974b). Lista alfabética dos vocábulos do questionário do inquérito linguístico e aditamentos posteriores. In M. de Paiva Boléo, *Estudos de Linguística Portuguesa e Românica*, vol.1, Dialectologia e História da Língua, t.2 (pp. 79–94). Universidade de Coimbra.
- Boléo, M. de Paiva (1974c). Mapa de Portugal Continental e Ilhas Adjacentes e lista das povoações nele inscritas. In M. de Paiva Boléo, *Estudos de Linguística Portuguesa e Românica*, vol.1, Dialectologia e História da Língua, t.2 (pp. 95–139). Universidade de Coimbra.
- Boléo, M. de Paiva (1978). *Inquérito Linguístico: Questionário*. 3.^a ed., Aveiro.
- Brissos, F. (2012). *Linguagem do Sueste da Beira no Tempo e no Espaço*. Centro de Linguística da Universidade de Lisboa.
- Brissos, F. (2016). Portugal: a Cidade e o Interior. I – Centro-sul. *Limite*, 10-1, 85–107. Disponível online em <http://www.revistalimite.es/volumen%2010/05brissos.pdf> (consulta: 16/01/2023).
- Brissos, F. (2018). Proposta de reformulação da classificação dialetal do noroeste português. *Estudos de Lingüística Galega*, vol. especial I, 193–208. <https://doi.org/10.15304/elg.ve1.3560>
- Brissos, F. (2022). Passado, presente e futuro da geografia linguística em Portugal. In I. Molina Martos & P. García Mouton (Eds.), *Geolingüística en la Península Ibérica* (pp. 113–136). Anejos de la Revista de Filología Española. CSIC.

- Cardoso, S., Mota, J., Aguilera, V., Aragão, M. do Socorro, Isquendo, A., Razky, A., Margotti, F., & Altenhofen, C. (2014). *Atlas Linguístico do Brasil*, vol.2, Cartas Linguísticas 1. EDUEL.
- Castro, I. (2016). Entrevista dada a Tânia Pinto Ribeiro para *Prelo*. Imprensa-Nacional Casa da Moeda. <https://impresanacional.pt/ivo-castro-em-entrevista/> (consulta: 16/01/2023).
- Cintra, L. F. Lindley (1971). Nova proposta de classificação dos dialectos galego-portugueses. *Boletim de Filologia*, 22, 81–116.
- Estrada Arráez, A., Benito Moreno, C. de (2016). Variación en las redes sociales: dados twilectales. *Revista Internacional de Lingüística Iberoamericana*, 14, 2(28), 77–111. <https://doi.org/10.31819/rili-2016-142806>
- García Mouton, P. (Coord.) (2009-). *ALPI-CSIC*. Edição digital de Tomás Navarro Tomás (Dir.), *Atlas Lingüístico de la Península Ibérica*. CSIC. <http://www.alpi.csic.es/> (consulta: 16/01/2023).
- Guardado, M. J., & Gomes, C. S. (2014). Evolução da população portuguesa. In M. Leston Bandeira (Dir.), *Dinâmicas Demográficas e Envelhecimento da População Portuguesa. 1950–2011: Evolução e Perspectivas* (pp. 29–109). Fundação Francisco Manuel dos Santos.
- Nascentes, A. (1953). *O Linguajar Carioca*. 2.^a edição completamente refundida. Organização Simões.
- Nunes, N., Rebelo, H., Saramago, J., & Vitorino, G. (2018). *Atlas Linguístico-Etnográfico da Madeira e do Porto Santo*, vol.1, A Criação de Gado: O Gado Bovino, Ovino e Caprino; O Leite e os Derivados; O porco e a Matança. Universidade da Madeira / Centro de Linguística da Universidade de Lisboa / Direcção Regional da Cultura.
- PORDATA – *Base de Dados Portugal Contemporâneo*. Fundação Francisco Manuel dos Santos. <http://www.pordata.pt/en/Home> (consulta: 16/01/2023).
- Projeto *Atlas Linguístico do Brasil (ALiB)*, página web: <https://alib.ufba.br/> (consulta: 16/01/2023).
- Romano, V. P., Seabra, R. D., & Oliveira, N. (2014). [SGVCLin] – Software para geração e visualização de cartas linguísticas. *Revista de Estudos da Linguagem*, 22-1, 119–151. <https://doi.org/10.17851/2237-2083.22.1.119-151>
- Saramago, J. (2006). O Atlas Linguístico-Etnográfico de Portugal e da Galiza (ALEPG). Separata de *Estudis Romànics*, 28, 281–298.
- Saramago, João (Coord.) (2012). *Atlas Linguístico-Etnográfico dos Açores*. Ed. digital, Direcção Regional da Cultura dos Açores. <http://www.culturacores.azores.gov.pt/alea/> (consulta: 16/01/2023).
- Vasconcelos, J. Leite de (1928). *Opúsculos*, vol. 2, Dialectologia (Parte 1). Coimbra: Imprensa da Universidade.
- Vasconcelos, J. Leite de (1929). *Opúsculos*, vol. 4, Filologia (Parte 2). Coimbra: Imprensa da Universidade.

- Vasconcelos, J. Leite de (1985). *Opúsculos*, vol. 6, Dialectologia (Parte 2). Lisboa: Imprensa Nacional-Casa da Moeda. Organizado por M. A. Valle Cintra.
- União Europeia (2021). *Path to the digital decade*. <https://digital-strategy.ec.europa.eu/en/library/proposal-decision-establishing-2030-policy-programme-path-digital-decade> (consulta: 16/01/2023).

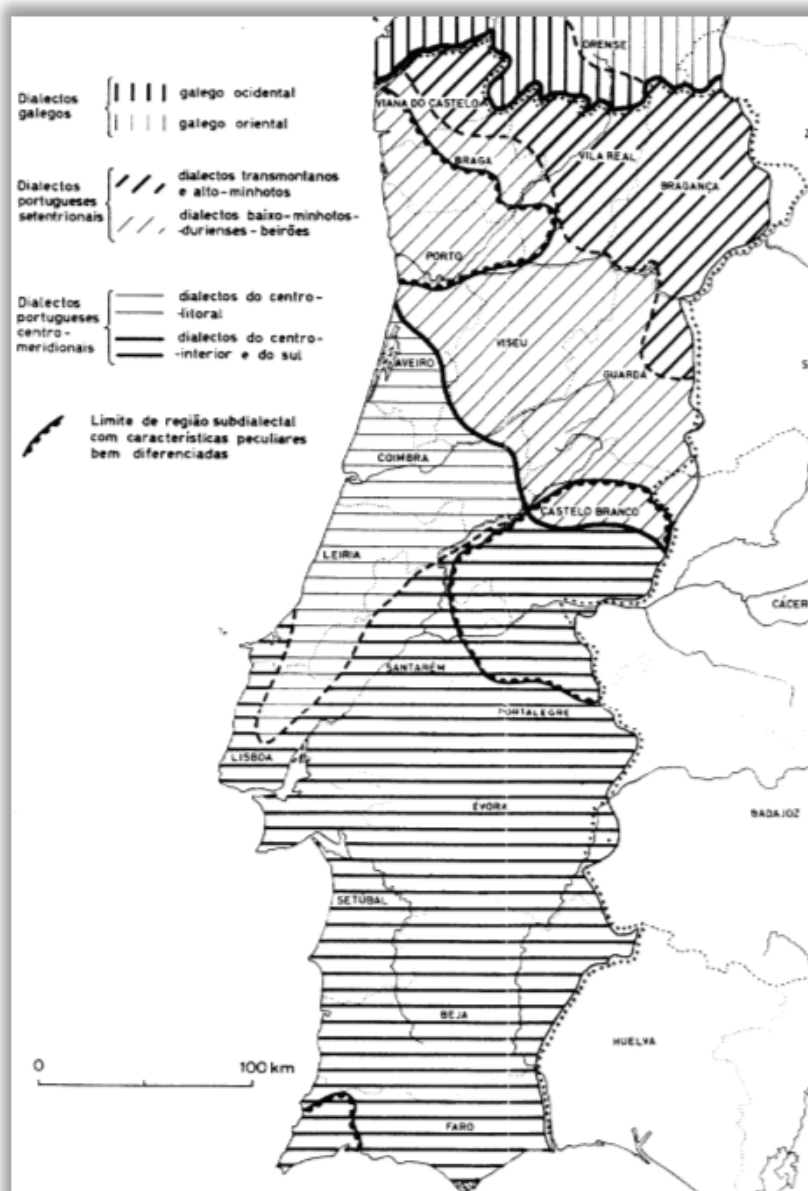
Recebido em: 25.01.2023

Aprovado em: 15.10.2023

EDITORA GERAL: Marcia Veirano Pinto 

Anexo

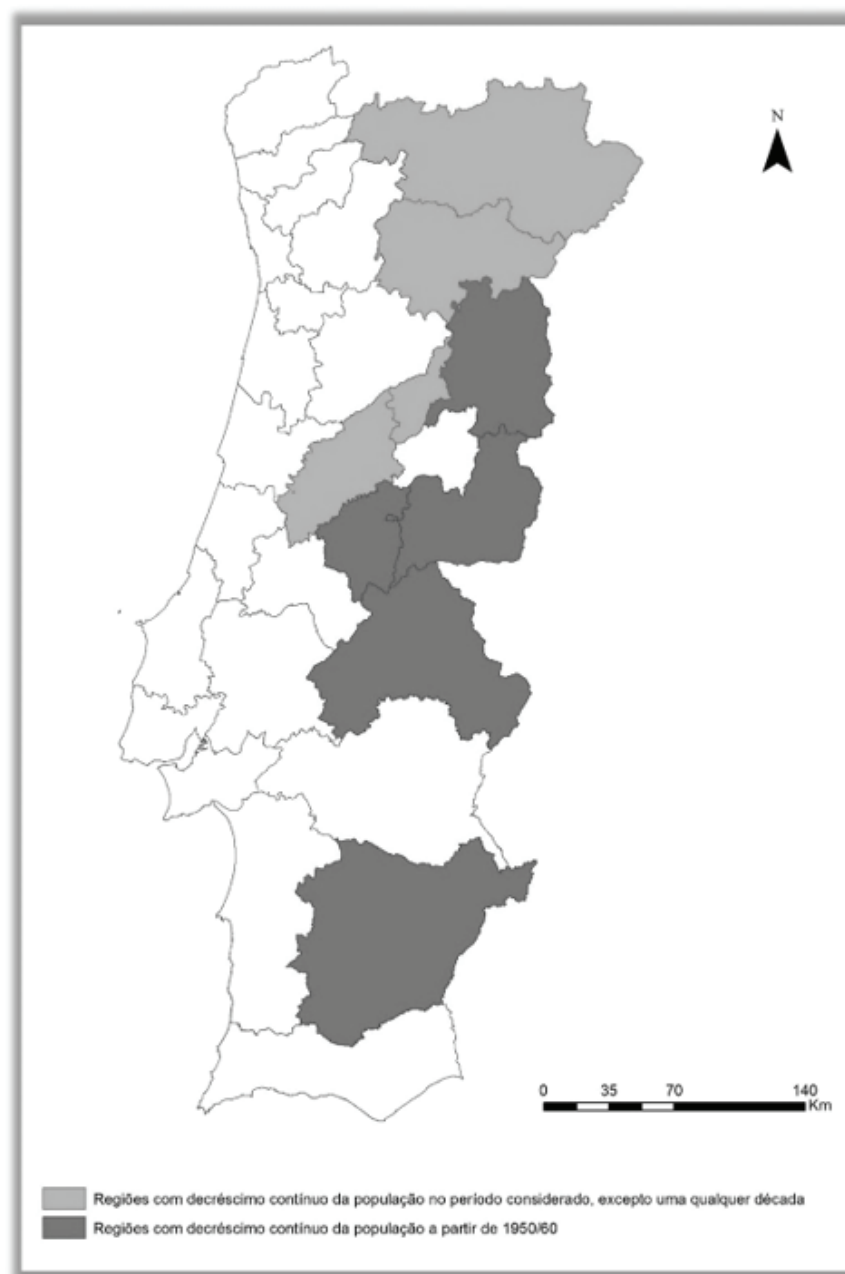
Mapa 1 — As áreas dialetais portuguesas continentais de acordo com Cintra (1971)



Mapa 2 — As áreas dialetais brasileiras de acordo com Antenor Nascentes (1953)



Mapa 3 — Regiões de Portugal continental com decréscimo populacional contínuo a partir de 1950, NUTS III.



Nota. Mapa elaborado a partir dos dados da Tabela 2; tanto o mapa como o quadro são retirados de Brissos (2016, p. 97-98).

Tabela 2 (anexa ao mapa 5) — Taxa de crescimento anual médio da população de Portugal continental a partir de 1950, NUTS III (%)

NUTS III	1950/1960	1960/1970	1970/1981	1981/1991	1991/2001	2001/2011
Minho Lima	0,12	-1,03	0,24	-0,26	0,01	-0,22
Cávado	0,76	0,2	1,71	0,71	1,08	0,43
Ave	1,52	0,34	1,52	0,76	0,91	0,03
Grande Porto	1,27	1,06	1,83	0,43	0,77	0,21
Tâmega	0,71	0,25	1,06	0,11	0,8	-0,01
Entre Douro e Vouga	1,32	1,12	1,16	0,63	0,94	-0,07
Douro	-0,21	-1,93	0,11	-0,9	-0,73	-0,74
Alto Trás-os-Montes	0,56	-2,58	-0,04	-1,45	-0,52	-0,88
Baixo Vouga	0,68	-0,22	1,51	0,4	0,97	0,13
Baixo Mondego	0,4	-0,39	1,14	-0,03	0,35	-0,24
Pinhal Litoral	0,73	-0,37	1,19	0,38	1,14	0,39
Pinhal Interior Norte	-0,69	-2,02	0,1	-0,86	-0,06	-0,52
Dão-Lafões	-0,04	-1,43	0,49	-0,43	0,14	-0,32
Pinhal Interior Sul	-0,33	-2,04	-1,59	-1,72	-1,26	-0,95
Serra da Estrela	-0,69	-1,9	0,05	-0,53	-0,8	-1,31
Beira Interior Norte	-0,82	-3,15	-0,44	-0,92	-0,27	-0,99
Beira Interior Sul	-0,42	-2,27	-0,85	-0,61	-0,37	-0,4
Cova da Beira	0,17	-2,16	-0,42	-0,69	0,05	-0,63
Oeste	0,33	-0,67	1,03	0,11	0,75	0,68
Médio Tejo	0,23	-0,75	0,4	-0,27	0,21	-0,24
Grande Lisboa	1,33	1,52	2,92	-0,09	0,35	0,48
Península de Setúbal	2,18	3,29	3,7	0,91	1,11	0,87
Alentejo Litoral	-0,13	-2,58	0,32	-0,45	0,15	-0,21
Alto Alentejo	-0,39	-2,54	-0,19	-0,58	-0,58	-0,7
Alentejo Central	-0,03	-2,08	0,11	-0,41	0,02	-0,4
Baixo Alentejo	-0,45	-3,03	-0,72	-1,04	-0,57	-0,64
Lezíria do Tejo	0,25	-0,66	1,02	-0,05	0,34	0,27
Algarve	-0,33	-1,6	1,85	0,53	1,49	1,33

Fonte: Guardado & Gomes (2014, p. 34-35).