

OS *CORPORA* DIALECTAIS DO CLUL: SUA CARACTERIZAÇÃO E OBJECTIVOS

Maria Luisa Segura da Cruz
CLUL

Ao apresentar os *corpora* dialectais que têm sido reunidos no Centro de Linguística da Universidade de Lisboa, através do Grupo de Estudos de Dialectologia, desde 1973, começarei por dizer que *corpus* é aqui tomado no seu sentido, mais lato e tradicional, de conjunto de dados reunidos com o objectivo preciso de servir determinada pesquisa e não no sentido mais restrito e actual – aquele em que tem sido considerado nesta mesa-redonda – de "colecção de textos de língua falada e/ou escrita, armazenada e processada em computador com o objectivo de ser utilizada em investigação linguística" segundo a definição de A. Renouf¹ e que alcança normalmente grandes dimensões.

Os *corpora* a que me vou referir inserem-se no âmbito da Geografia Linguística e visam estudar, do ponto de vista fonético, fonológico, morfo-sintáctico e lexical, a variação regional existente no território português, considerado no seu todo ou em parcelas mais ou menos circunscritas desse território. O seu objectivo final é permitir a elaboração de atlas linguísticos ou linguístico-etnográficos nacionais ou regionais que revelem, de modo sistemático, essa variação.

A constituição de *corpora* em Geografia Linguística é, pois ditada e está condicionada pelos pressupostos científicos e pelos objectivos desta disciplina, pelas metodologias e técnicas de apresentação de materiais que lhe são próprias.

Pretendendo a Geografia Linguística descrever e analisar a diversidade dialectal, o *corpus* de estudo deverá ser concebido de modo a permitir, por um lado caracterizar linguisticamente cada falar e, por

outro, comparar os diferentes falares entre si. A uniformidade de critérios na escolha de localidades e de informantes e a homogeneidade dos temas e conceitos abordados em todo o território asseguram a possibilidade de comparar os dados. Só assim se tornará possível conhecer a extensão geográfica de determinados fenómenos, traçar isoglossas, caracterizar e delimitar áreas e sub-áreas dialectais, confrontar e interpretar correspondências entre fenómenos geograficamente afastados, descobrir eventuais sobrevivências de fenómenos já desaparecidos na língua padrão, aspectos que, entre outros, se espera sejam revelados por um atlas linguístico.

Deste modo, quanto às localidades, as recolhas processam-se em comunidades rurais ou piscatórias pouco populosas, variando o seu número consoante o tipo de atlas a que se destina (atlas nacional ou regional). Por sua vez, a população inquirida é constituída por indivíduos considerados representativos das respectivas comunidades, de ambos os sexos, de idade não inferior a 50 anos, preferentemente analfabetos ou de reduzida escolaridade, com poucos ou nenhuns contactos com meios exteriores, de modo a reduzir a possibilidade de interferências de normas alheias à sua, nomeadamente a influência da escolaridade. Quanto à homogeneidade dos dados linguísticos colectados, ela é proporcionada e garantida pela utilização de um questionário aplicado em todos os pontos da rede.

Os *corpora* dialectais que se têm constituído ou cuja constituição ainda está em curso têm, pois, as seguintes características em comum:

1. o mesmo tipo de população,
2. o mesmo tipo de lugar de recolha,
3. serem *corpora* orais,
4. serem integralmente gravados.

Quanto a este aspecto, aproveito para referir que o arquivo sonoro constituído pelos vários *corpora* reunidos e que pode ser consultado por quem estiver interessado no estudo dos dialectos ultrapassa já as 3000 horas de gravação.

Quanto à sua constituição, os *corpora* são de dois tipos:

1. *corpus* obtido a partir das respostas a um questionário previamente estabelecido, o mais significativo.
2. *corpus* de conversas livres, ou seja, de texto oral espontâneo.

Passarei agora à apresentação de cada um dos *corpora*:

1. Atlas linguístico-etnográfico de Portugal e da Galiza (ALEPG)

Criado por Lindley Cintra em 1970 e de que é actualmente responsável João Saramago, o ALEPG é um projecto de grande envergadura, cujo objectivo é a publicação de um atlas nacional. Circunstâncias várias, entre as quais se pode apontar a exiguidade de verbas e de pessoal afectos a uma empresa de tamanho vulto, têm retardado a conclusão do projecto que conhece, neste momento, uma fase de intensificação das recolhas².

As recolhas para este atlas iniciaram-se em 1973 e prevê-se que estejam concluídas (pelo menos a sua primeira fase) em 1996.

A rede inicialmente prevista contemplava cerca de 300 pontos de inquérito no que diz respeito a Portugal continental e insular e foi finalmente fixada em 214, com a seguinte distribuição:

Portugal continental: 178 localidades,
Madeira e Porto Santo: 7 localidades,
Açores: 17 localidades,
Pontos fronteiriços: 12 localidades em território politicamente espanhol.

O questionário utilizado está organizado em campos semânticos, segundo a divisão proposta por Hallig e von Wartburg³o que permite um encadeamento lógico das perguntas e foi concebido de modo a que incidindo sobre o léxico (abordado numa perspectiva onomasiológica), se obtenha a informação necessária à descrição fonética, fonológica e de aspectos importantes da morfologia de cada falar. Por exemplo para a caracterização fonológica foram incluídas sistematicamente questões cujas respostas correspondem a pares mínimos que dêem conta de oposições fonológicas. Para o estudo da morfo-fonologia, incluíram-se, por exemplo, perguntas cujas respostas ilustrem a flexão de género e número, aspectos relativos à metafonía nominal e verbal ou à flexão verbal.

O questionário está dividido em capítulos incidindo sobre uma grande variedade de temas, dos quais destacamos, a título de exemplo:

- o céu e a atmosfera,
- a terra e a configuração do terreno,
- as plantas (árvores, ervas e flores),
- os animais (mamíferos, aves, insectos e outros invertebrados, batráquios e répteis),
- o homem e o trabalho (cultivo dos campos, aproveitamento dos produtos vegetais, criação de gado, ofícios e tecnologias tradicionais).

Este questionário, constituído por cerca de 3500 perguntas, foi aplicado, até 1990, em 71 localidades. Pela necessidade de fazer, o mais brevemente possível, a cobertura total do país, para possibilitar o tratamento de dados à escala nacional e iniciar a publicação do atlas, houve a necessidade de reduzir o número de perguntas para cerca de metade, o que encurta consideravelmente o tempo de recolha, remetendo as restantes para uma segunda etapa.

O questionário reduzido foi já aplicado em 89 pontos incluindo a totalidade de pontos da Madeira e Porto Santo, e três nos Açores, esperando-se que a primeira fase de recolhas esteja terminada no próximo ano.

Para além das respostas obtidas pela aplicação do questionário, e para obviar às limitações que o sistema pergunta/resposta representa na produção de textos orais, lugar privilegiado para a observação e análise de fenómenos da sintaxe dialectal, procede-se sistematicamente à recolha de conversas livres em que os inquiridos descrevem actividades tradicionais, acontecimentos da vida pessoal e local, contam episódios, histórias, etc. Estas narrativas permitem, também, esclarecer aspectos etnolinguísticos relacionados com os próprios temas em estudo. Com menor sistematicidade, mas sempre que possível, recolhe-se literatura oral: romances, rezas, orações, contos, fórmulas infantis, décimas, quadras, adivinhas, provérbios.

O *corpus* de respostas para este projecto serve igualmente o projecto internacional **Atlas Linguístico Românico (ALIR)**⁴ no qual Portugal está representado através do Grupo de Estudos de Dialectologia do CLUL.

2. Atlas linguístico do litoral português (ALLP):

Este projecto, da responsabilidade de Gabriela Vitorino, tem por objectivo o estudo da linguagem específica dos pescadores do litoral português, incidindo apenas sobre o léxico.

O questionário utilizado contém cerca de 1200 perguntas, distribuídas pelos seguintes capítulos:

1. a pesca e os processos de pesca,
2. embarcações e navegação,
3. a companhia e a comercialização do pescado,
4. o peixe (generalidades),
5. a fauna e a flora marinhas,
6. a costa e o mar,
7. fenómenos atmosféricos.

A rede de pontos é constituída por 23 localidades (portos de pesca) do litoral continental português.

O projecto tem-se desenvolvido em várias fases, estando concluída a primeira, que deu origem à elaboração de um primeiro volume consagrado à fauna e à flora marinhas e que inclui também o capítulo 4. O referido volume, que aguarda publicação, contém 200 mapas e respectivos comentários e está redigido em dois tomos⁵.

O questionário completo foi já aplicado em 8 dos 23 pontos da rede. Posteriormente à elaboração do primeiro volume, foram feitas recolhas referentes à fauna e flora marinhas em 4 pontos do arquipélago da Madeira e 2 dos Açores (Faial e Pico), estando previstas ainda em 1995 recolhas em mais 2 ilhas açorianas (Flores e São Jorge).

3. Atlas linguarum europae (ALE):

Curiosamente o primeiro *corpus* que se encontra integralmente recolhido destina-se a um projecto internacional, o ALE⁶, que, como é sabido, recobre todo o continente europeu, do Atlântico aos Urais e ao Cáucaso.

As recolhas tiveram lugar em 1975, tendo sido utilizado um questionário de 546 perguntas de interesse exclusivamente lexical. Foi aplicado em 53 localidades portuguesas do continente.

Foram publicados até agora 4 volumes de mapas interpretativos, acompanhados de outros tantos volumes de comentários⁷, estando o quinto no prelo.

4. Barlavento do Algarve (BA):

Trata-se de um *corpus* reunido com o objectivo de estudar exclusivamente aspectos fonéticos (vocalismo) de uma única variedade dialectal, o Barlavento do Algarve, no intuito de traçar as suas fronteiras⁸.

Esse *corpus*, recolhido por mim própria, em 1986 e 1987, é constituído pelas respostas a um questionário de apenas 378 questões. Três critérios presidiram à elaboração do questionário: 1) a inclusão de palavras que ilustrassem a evolução das vogais do latim para o português, em sílaba acentuada e não acentuada em posição inicial, medial e final, em sílaba livre e travada. A escolha de palavras em função deste critério principal, teve ainda em linha de conta outros dois factores: 2) a utilização de palavras correntes que designassem realidades conhecidas dos informantes; 3) a inclusão de palavras para as quais se obtivessem respostas sem variação lexical.

Este questionário foi aplicado integral ou parcialmente em 53 localidades do BA.

Nas mesmas localidades foram também gravados textos livres, versando, dum maneira geral temas fixos: o cultivo da terra, técnicas e instrumentos de trabalho (o arado, o moinho, a mó manual, o fabrico do pão), actividades piscatórias nas localidades do litoral.

5. O falar da ilha do Corvo

Este projecto, da responsabilidade de João Saramago, estudou o dialecto corvino com um duplo objectivo: realizar um estudo fonético, do ponto de vista acústico e um estudo lexical numa perspectiva socio-linguística.⁹

Foram usados dois questionários:

- um questionário lexical abrangendo 908 conceitos, aplicado a 18 locutores (9 homens e 9 mulheres), divididos por escalões etários: 3 homens e 3 mulheres da faixa etária dos 25 – 30 anos; 3 homens e 3 mulheres com idades compreendidas entre os 45 e os 50 anos e 3 homens e 3 mulheres com idades superiores a 70 anos.

A aplicação deste questionário visou: a) a análise e classificação do léxico obtido de acordo com a dicotomia dialectal / não dialectal; b) a análise da variação lexical interna, segundo a dicotomia acima

referida e de acordo com: 1) o sexo e idade dos informantes; 2) a idade dos informantes; 3) o sexo dos informantes.

• um questionário fonético, aplicado a 3 locutores masculinos (um de cada escalão etário) destinado a estudar, do ponto de vista acústico, as vogais orais, acentuadas e não acentuadas, do dialecto corvino.

Informatização dos materiais

Só muito recentemente se iniciou a informatização de alguns destes materiais dialectais. Em 1994, foi criada uma base de dados e o software de apoio à informatização dos volumosos materiais recolhidos no âmbito do projecto ALEPG.

O desenho da base de dados virá a permitir:

– realizar a cartografagem automática dos dados, o que representará qualitativamente, um avanço qualitativo substancial no processo de apresentação, no rigor e na fiabilidade dos mapas, relativamente à programação inicial.

– tornar acessíveis à consulta uma multiplicidade de aspectos, de que se destacam entre outras possibilidades: a de conhecer total ou parcialmente as respostas de cada ponto da rede, o que permitirá caracterizar mais facilmente qualquer falar; conhecer em cada ponto da rede as respostas de cada informante; conhecer, para cada conceito, as respostas obtidas, na totalidade ou em parte das localidades inquiridas; conhecer a extensão e distribuição geográfica de qualquer fenómeno fonético, ou delimitar geograficamente áreas lexicais.

Resta dizer que estamos ainda no início desta fase dos trabalhos, sendo o primeiro objectivo a informatização do *corpus* relativo às recolhas açorianas, já iniciada, e que darão origem à primeira publicação do ALEPG.

Notas

¹ Antoinette Renouf, 1987, "Corpus Development", in Sinclair, J. M. (ed.) *Looking Up. An account of the CoBUILD Project in lexical computing*. London, Collins CoBUILD: 1-42, p.1

² O ALEPG está temporariamente cindido em dois projectos: *Corpus para um Atlas linguístico de Portugal* e *Atlas linguístico-etnográfico dos Açores*, que contam

com o apoio financeiro respectivamente dos Programas Lusitânia e Estímulo da JNICT.

- ³ R. Hallig e von Wartburg, 1952, *Begriffssystem als Grundlage für die Lexikographie*, Berlim.
- ⁴ Este projecto, criado em 1987, congrega todos os países de língua românica e tem sede em Grenoble.
- ⁵ Gabriela Vitorino, 1988, *Atlas linguístico do litoral português. Fauna e flora.* Dissertação apresentada para progressão na carreira de investigação. INIC-CLUL.(inédita).
- ⁶ O ALE foi criado por iniciativa da Universidade Católica de Nijmegen (Holanda), sob os auspícios da UNESCO e nele participam todos os países europeus, tendo actualmente a sua sede em Florença.
- ⁷ *Atlas Linguarum Europae*, Mario Alinei et alii, T. I- IV (1983-1990), Van Gorcum, Assen.
- ⁸ Maria Luisa Segura da Cruz, 1988, *A fronteira dialectal do Barlavento do Algarve.* Dissertação para progressão na carreira de investigação. INIC - CLUL. (no prelo).
- ⁹ João Saramago, 1992, *Le parler de l'île de Corvo, Açores.* Centre de Dialectologie, Université de Stendhal, Grenoble III, Grenoble.