

Arquivo Dialectal do CLUP

Tratamento linguístico e cartográfico de um
conjunto de
amostras dialetais do português europeu:
primeiros resultados

U. PORTO

ARQUIVO
DIALETAL
DO CLUP

CLUP Centro de
Linguística da
Universidade do
Porto

João Veloso^{1,2} e Pedro Tiago Martins^{2,3}

¹Faculdade de Letras da Universidade do Porto

²Centro de Linguística da Universidade do Porto

³Universidade de Barcelona

Breve apresentação do Arquivo Dialetal

- Coleção de amostras dialetais do português europeu contemporâneo
- Resultado de recolhas informais efetuadas desde 1994
- **Objetivo:** criação de um recurso on-line, disponível em acesso aberto a qualquer interessado
- Cada amostra:
 - ficheiro sonoro;
 - transcrição ortográfica e transcrição fonética estreita;
 - informação sociodemográfica do informante;
 - informação cartográfica.
 - levantamento de principais marcas dialetais

História do projeto: Início

- 1994, como elemento (opcional/obrigatório) para avaliação curricular em várias disciplinas de Linguística Portuguesa na FLUP
- Recolha de ca. 1-2 min. de fala.

História do projeto: Início

- Indicações gerais dadas aos estudantes:
 - Variedade diferente da do observador, sempre que possível;
 - Entrevista espontânea, de duração superior a 1 min;
 - Seleção, dentro da amostra total, da porção mais rica em termos de variação dialetal;
 - Elementos a entregar:
 - Ficheiro sonoro, transcrição ortográfica, transcrição fonética, caracterização do falante, comentário dialetal, anexos e bibliografia.
- Sem especificações muito estritas relativamente à qualidade e/ou ao suporte da gravação.

História do projeto: Início

- Única motivação inicial: estimular junto dos estudantes a recolha de dados, o trabalho prático e de campo (“trabalho escolar”)
- **Vantagens:** atraiu muitos estudantes e permitiu reunir uma vasta coleção de material dialetal.
- **Desvantagens:** grande parte das recolhas (e do seu tratamento) apresenta uma deficiente qualidade sonora ou de análise linguística.

Crescimento desorganizado de todo o acervo

- Material depositado em várias localizações diferentes;
- Material perdido;
- Algumas das recolhas faziam parte de trabalhos maiores.
- Diferentes elementos de algumas das recolhas em falta.

História do projeto: Início formal

- 2008: Bolsas BII da FCT:
 - 2008/2009: 2 bolseiros
 - 2009/2010: 1 bolseiro
- ✓ **Catálogo** de todas as amostras
- ✓ **Revisão criteriosa de toda a informação demográfica e linguística**
- ✓ **Levantamento da informação linguística e demográfica**

- Base de dados fiável de todo o material, com informação linguística relevante e caracterização completa de aspetos como:
 - N° de amostras
 - Data das amostras
 - Distribuição dos inquiridos por idade e origem geográfica
 - ...
- Sugestão de algum potencial para a ilustração e a investigação de aspetos da variação dialetal

História do projeto: Rearranque

- 2011/2012:

Financiamento CLUP → 1 investigador a tempo parcial:

- ✓ revisão geral da catalogação BII;
- ✓ verificação criteriosa de toda a informação geográfica de cada amostra;
- ✓ reconstrução da base de dados original (1ª fase do projeto)
- ✓ transcrição ortográfica e transcrição fonética estreita de todas as amostras (ignorando completamente as transcrições que acompanhavam os trabalhos dos estudantes);
- ✓ seleção das amostras com maior qualidade;
- ✓ criação de página internet.

Seleção das amostras

Objectivo: Selecionar as amostras com maior qualidade para uma descrição linguística completa e inserção na BD da futura página web.

Critérios:

- Informação sociodemográfica completa;
- Qualidade técnica e inteligibilidade da gravação.
- Verificação e omissão de dados que pudessem revelar ou a partir dos quais fosse possível inferir a identificação civil dos informantes.

Descrição geral do acervo AD-CLUP:

Coleção de todas as amostras recolhidas (1994-2012)

- TOTAL de Amostras = 141

Informantes do sexo masculino: 56

Informantes do sexo feminino: 85

Proveniência geográfica dos informantes

(por distrito/região autónoma) – COLEÇÃO COMPLETA DO ARQUIVO

Aveiro	14	Açores	5
Braga	11	Madeira	7
Bragança	5		
Coimbra	3		
Faro	4		
Guarda	1		
Leiria	4	Distritos sem amostras recolhidas:	Castelo Branco, Évora, Beja
Lisboa	7		
Portalegre	1		
Porto	49		
Santarém	1		
Setúbal	1		
Viana do Castelo	6		
Vila Real	9		
Viseu	6		
Desconhecido	7		

Descrição geral do acervo AD-CLUP:

Subconjunto das amostras a disponibilizar on-line

- TOTAL de Amostras = 71

Informantes do sexo masculino: 34

Informantes do sexo feminino: 37

Proveniência geográfica dos informantes

(por distrito/região autónoma) – AMOSTRAS SELECIONADAS

Aveiro	7	Açores	2
Braga	9	Madeira	4
Bragança	3		
Coimbra	3		
Faro	3		
Guarda	1		
Leiria	2	Distritos sem amostras recolhidas:	Castelo Branco, Évora, Beja
Lisboa	5		
Portalegre	1		
Porto	18		
Santarém	---		
Setúbal	---		
Viana do Castelo	4		
Vila Real	3		
Viseu	6		
Desconhecido	---		

Algumas limitações

- formatos analógicos (cassete, mini-cassete, etc.)
- qualidade técnica/inteligibilidade de algumas gravações abaixo do desejável
- desnível na qualidade dos trabalhos
- informalidade e disparidade nos critérios da recolha

Resultados do projeto

- Organização de toda a coleção de amostras
- Análise linguística de dados
- Interesse por parte de investigadores de outras instituições (Brasil, Finlândia)
- Investigação sobre o próprio Arquivo (transcrição fonética, cartografia de dados dialetais)
- Melhoramento /Uniformização das condições e especificações técnicas das futuras recolhas
- Página web

A página *web*

- Grande objetivo do Arquivo
- Recurso on-line, aberto a qualquer interessado
- Público-alvo: estudantes e professores de Português/Linguística Portuguesa e Geral; investigadores na área da variação dialetal do português (com reservas)

Informação Disponibilizada na Página

- Dados gerais do Arquivo (história, descrição, etc.);
- Documentação técnica acerca do Arquivo e do seu material;
- Para cada amostra:
 - Gravação sonora (wav);
 - Informação sociodemográfica;
 - Transcrição ortográfica (pdf);
 - Transcrição fonética (pdf);
 - Transcrição alinhada (pdf)
 - Informação cartográfica;
 - Informação linguística.
- Dados estatísticos;
- Mapas gerais e de cruzamento de variáveis linguísticas;
- Motor de busca (futuramente);
- Lista de autores das recolhas.



Amostras não incluídas na página

- **Motivos:**
 - Má qualidade das gravações;
 - Dificuldade de transcrição/levantamento de marcas dialetais;
 - Gravações em formatos pouco comuns ainda em fase de conversão para formatos digitais
 - Dados importantes em falta
 - Problemas técnicos alheios ao Arquivo
 - Identificação civil óbvia dos informantes
- **Acesso:**
 - Investigadores interessados podem solicitar a consulta das amostras não incluídas na página.

Variáveis linguísticas

- Um dos contributos originais do projeto: estudo direcionado para algumas variações específicas.
 - Distinção /b/~v/
 - Realização de róticos: [r], [R], [ʁ], [x], [χ]
 - Realização/apagamento de chevá
 - Coarticulação de [ɐ] com outras vogais: [o^ɐ], [e^ɐ], [ɛ^ɐ] (transcrição proposta)

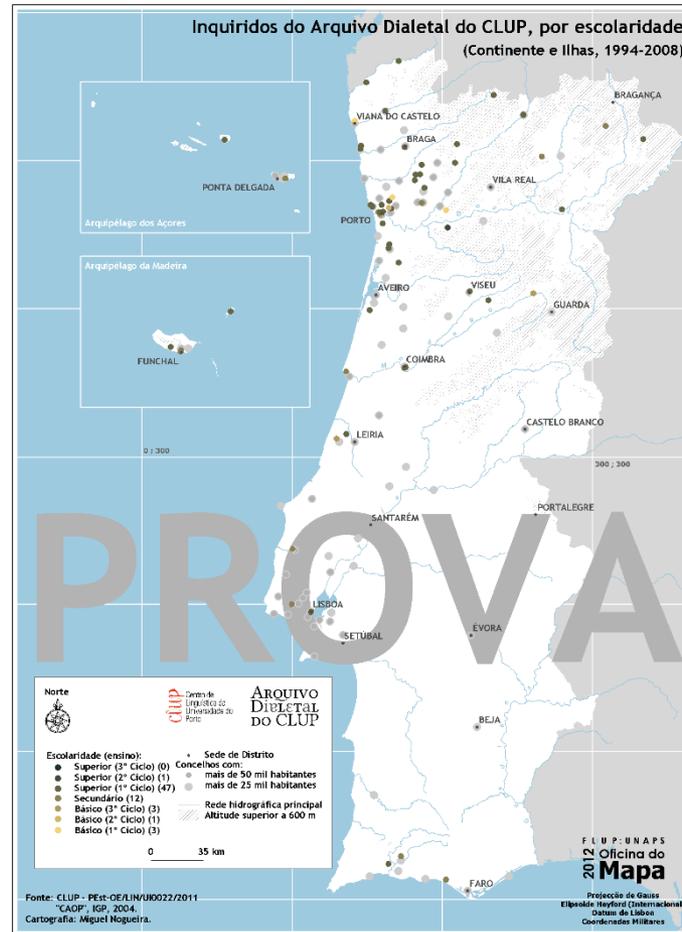
Convenções e validação das transcrições fonéticas

- **Outro contributo:** proposta de algumas convenções fonéticas
 - Ex^o: [o^v], [e^v], [ɛ^v]
- Validação das transcrições fonéticas
 - Várias fases de revisão;
 - Análise acústica em casos de dúvida (*PRAAT*)
 - "Inter-judge Agreement" (Martins & Veloso 2012)

A cartografia neste projeto

- Recurso não apenas ilustrativo. Parte integrante do desenvolvimento do próprio projeto (Oficina do Mapa, FLUP);
- Método intuitivo de apreensão do conteúdo do Arquivo;
- Facilidade na detecção de padrões;
- Grande qualidade tipográfica.

Exemplo de cartografia: distribuição geográfica dos inquiridos (amostras selecionadas) por grau de escolaridade (versão provisória)



Exemplo de um dos ficheiros do Arquivo

- [27_A.pdf](#)



Agradecimentos

- Pessoas:

- Marta Afonso (Centro de Linguística da Universidade do Porto)
- Ana Paula Alencastro (Pontifícia Universidade Católica do Rio Grande do Sul)
- Rita Ferreira (Faculdade de Letras da Universidade do Porto)
- Diamantino Freitas (Faculdade de Engenharia da Universidade do Porto)
- Manuel Gericota (Instituto Superior de Engenharia do Porto)
- Luís M. T. Jesus (Instituto de Engenharia Electrónica e Telemática de Aveiro)
- Ilpo Kempas (Universidade de Seinäjoki)
- Miguel Nogueira (Oficina do Mapa, Universidade do Porto)
- Catarina Oliveira (Instituto de Engenharia Electrónica e Telemática de Aveiro)
- Margarida Gil Pires (Faculdade de Letras da Universidade do Porto)
- Iiris RENNICKE (Universidade de Helsínquia)
- Autores de todas as recolhas (Faculdade de Letras da Universidade do Porto)

- Instituições:

- Faculdade de Letras da Universidade do Porto
- Centro de Linguística da Universidade do Porto
- Oficina do Mapa (FLUP)
- Departamento de Engenharia Electrotécnica e de Computadores da Universidade do Porto
- Fundação para a Ciência e a Tecnologia

Referências

- Martins, P. T.; Veloso, J. 2012. *Inter-judge Agreement in Transcribing Dialectal Data: A Study of a Corpus of Dialectal Portuguese*. Poster apresentado na “45th Annual Meeting of the Societas Linguistica Europaea”, Universidade de Estocolmo, Suécia.