CreoPhonPt: introdução à prática científica na era digital com impacto sobre a preservação dos crioulos

Carlos Silva[1]
Luís Trigo[2]
Vera Moitinho de Almeida[3]

Resumo

O CreoPhonPt (Silva & Moran, 2022) é um repositório de acesso livre que reúne vários *datasets* com informação fonológica, lexical e sociolinguística dos crioulos do português. Este repositório foi criado no âmbito de um projeto de doutoramento, com o intuito de facilitar a análise quantitativa destas línguas, tornando, em simultâneo, acessíveis, dados que anteriormente estavam fechados e/ou dispersos. Um novo ramo do projeto – CreoPhonPt_classes – surgiu quando o seu fundador assumiu a docência da unidade curricular (UC) de Fonologia I do mestrado em Linguística da FLUP. Este projeto de seguimento teve base em dois grandes objetivos imediatos:

[1] FCUP. *Email*: carlos.silva@fc.up.pt
[2] FMUP. *Email*: ltrigo@letras.up.pt
[3] FLUP. *Email*: vmoitinho@letras.up.pt

1. Enriquecer o repositório existente com mais dados, potenciando o seu uso e divulgação;
2. Facultar aos estudantes as capacidades de:
   2.1. recolha estruturada de dados científicos;
   2.2. acompanhamento rigoroso de protocolos de investigação;
   2.3. processamento manual e automático dos dados em folhas de cálculo e em código Python;
   2.4. ligação dos dados e metadados a redes abertas, uniformização e divulgação.

As aulas, lecionadas em parceria com investigadores de diferentes áreas das humanidades digitais, tornaram-se numa interessante experiência de interdisciplinaridade e de colaboração, cujos resultados positivos se manifestaram tanto no bom aproveitamento dos estudantes, como nas classificações atribuídas à UC nos inquéritos pedagógicos.

Abstract

CreoPhonPt (Silva & Moran, 2022) is an open-access repository that combines various datasets with phonological, lexical and sociolinguistic information on the creoles of Portuguese. This repository was created as part of a PhD project to facilitate the quantitative analysis of these languages while at the same time making accessible data that was previously closed

and/or dispersed. A new branch of the project – CreoPhonPt_classes – emerged when its founder took over the teaching of Phonology I on the MA in Linguistics at FLUP. This follow-up project was based on two main immediate objectives:

1. Enrich the existing repository with more data, boosting its use and dissemination;
2. Provide students with the skills to:
   2.1. structured collection of scientific data
   2.2. rigorous monitoring of research protocols;
   2.3. manual and automatic data processing in spreadsheets and Python code;
   2.4. linking data and metadata to open networks, standardisation and dissemination.

The classes, taught in partnership with researchers from different areas of the digital humanities, became an exciting experience of interdisciplinarity and collaboration, the positive results of which were evident both in the excellent performance of the students and in the ratings given to the UC in the pedagogical surveys.

Palavras-chave
CreoPhonPt; Crioulo do português; Repositório.

Keywords
CreoPhonPt; Portuguese creole; Repository.

Introdução

A Fonologia dos crioulos de base portuguesa (bem como de outras línguas) é um campo da Linguística ainda pouco explorado. Por um lado, os linguistas que fazem a recolha de dados no campo não são, na sua maioria, fonólogos ou foneticistas. Consequentemente, as descrições dos crioulos do português relatam pouco mais do que o inventário segmental destas línguas em termos dos seus sistemas de sons. Por outro lado, a fonologia teórica – sistemas de traços, estruturas suprassegmentais, etc. – tem por base o estudo das línguas europeias, o que leva os fonólogos a concentrarem-se também nestes para corroborar ou rebater propostas anteriores. Para além disso, nos estudos de fonologia histórica, as línguas de contacto, como os crioulos e os pidgins, são muitas vezes evitados dada a ausência de classificação genética (Wichmann & Holman, 2009) e o elevado número de empréstimos que incorporam e que apresentam irregularidades acrescidas (Greenhill *et al.*, 2017).

Estas razões, bem como o facto de muitos crioulos de base portuguesa serem línguas em vias de extinção, tornam urgente, em primeiro lugar, a recolha sistemática de dados fonológicos sólidos e a composição de uma base de dados localizável, acessível, interoperável e reutilizável. Em segundo lugar, urge a necessidade de retomar esses dados em estudos fonológicos que descrevam e deem visibilidade à chamada "Fonologia dos crioulos". De facto, estes dados e o seu

processamento podem ser usados de forma vital na preservação de amplo património linguístico.

Contexto científico da prática pedagógica

O projeto CreoPhonPt, designadamente o ramo CreoPhonPt_classes, foi desenvolvido no âmbito da UC de Fonologia I do mestrado em Linguística da Faculdade de Letras da Universidade do Porto. Nesta UC, pretende-se que os estudantes não só adquiram conhecimentos inerentes à teoria fonológica, mas também sejam capazes de os aplicar mediante a descrição das línguas naturais. Este segundo objetivo é difícil de aplicar a línguas como português e outras línguas europeias, na medida em que estas já se encontram amplamente descritas por diferentes correntes teóricas. Foram, por isso, selecionados como campo de aplicação os crioulos do português, cuja lacuna bibliográfica no que diz respeito à Fonologia e suas componentes (fonema, sílaba, acento, etc.) constitui *per se* uma motivação para os estudantes.

Estratégias pedagógicas utilizadas

Esta UC totaliza 30 horas letivas divididas em 10 blocos de 3 horas cada. Uma das estratégias usadas para, por um lado, aplicar o conceito de aprendizagem baseada em

projeto (Li *et al.*, 2015) e, por outro, prevenir o déficit de atenção dos estudantes, foi a divisão de cada um destes blocos em duas partes: 1h45 de exposição teórica e 1h de trabalho prático. Em cada sessão, as duas partes estavam relacionadas. Por exemplo, na aula em que foi exposto e discutido o conceito de fonema na parte teórica, os estudantes tiveram oportunidade, na parte prática, de reconstruir criticamente o inventário de um crioulo do português escolhido pelos mesmos. Da mesma forma, na aula acerca da sílaba, os estudantes foram acompanhados no processamento dos dados silábicos do crioulo previamente selecionado por eles.

Os dados foram processados individualmente por cada estudante, mas acompanhados em permanência pela equipa de docentes. A estruturação e processamento manual dos dados foi realizada com recurso à folha de cálculo Google Sheets. Para o processamento automático, quer através de fórmulas de cálculo, quer através de código Python, utilizou-se o Google Sheets e a aplicação de *notebooks* do Google Colab. A limpeza dos dados e a ligação a dados abertos foi feita através da plataforma OpenRefine. Finalmente, o repositório com controlo de versões GitHub serviu para a criação de uma organização composta pelos elementos da turma e pelos docentes e de um repositório de livre acesso com controlo de versões.

Todas as ferramentas apresentadas são de acesso gratuito (de código aberto e/ou online) e fazem parte de um leque indispensável ao tratamento de dados, tanto noutras áreas da Academia, quanto na indústria.

Inovação pedagógica

É típico, em várias UC do mestrado, que o docente avalie os estudantes através de um trabalho escrito, escolhido livremente por eles ou proposto pelo docente. Esses trabalhos, depois da avaliação, são na maioria dos casos arquivados, ou seja, os estudantes não os continuam a desenvolver, nem os dados/resultados são retomados por outros estudantes ou investigadores. Eles ficam, portanto, na esfera individual.

A inovação que se procurou com este tipo de projeto foi um acompanhamento mais próximo dos estudantes, introduzindo-os na prática real de investigação num ambiente colaborativo e aberto. O foco, na parte metodológica, com o ensino de processamento computacional dos dados, faz com que a prática pedagógica nesta UC se distinga de outras que se concentram quase exclusivamente na transmissão e discussão de conteúdos teóricos. Procura-se também transmitir aos estudantes a importância da transparência, da repetibilidade e reprodutibilidade e da continuidade da investigação, que deve ser, não um produto destinado ao investigador, mas ao benefício da sociedade em geral.

Resultados esperados

Resultados obtidos para o caso de unidades curriculares existentes, nomeadamente indicadores de desempenho dos estudantes e/ou análise aos resultados dos inquéritos pedagógicos.

A exposição dos trabalhos dos estudantes num repositório de livre acesso, bem como o treino no tratamento de dados resultou numa maior motivação por parte dos estudantes e na sua satisfação com a UC, como comprovam os resultados obtidos nos inquéritos pedagógicos:

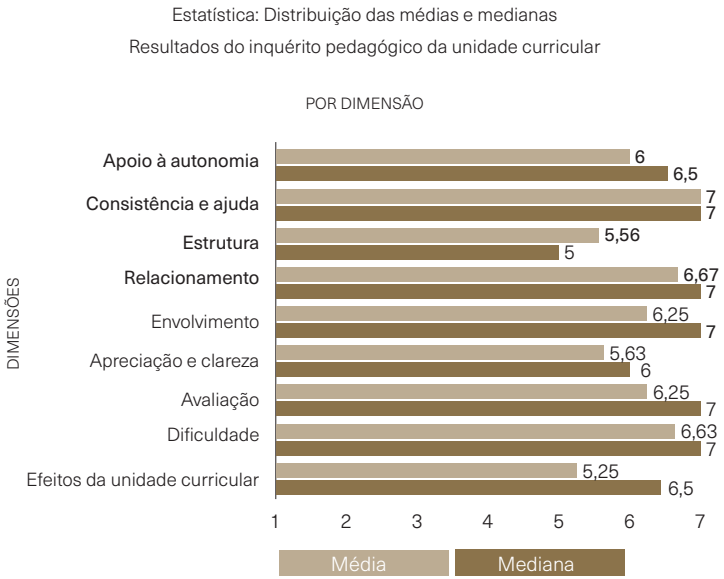

FIGURA 1 • Resultados do inquérito pedagógico da unidade curricular.

Referências

Greenhill, S., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C. & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences 114(42)*. E8822–E8829. DOI: 10.1073/pnas.1700388114. Available at: https://www.pnas.org/content/114/42/E8822.

Li, D., Zhang, C., & He, Y. (2015). Project-based learning in teaching translation: students' perceptions. *The Interpreter and Translator Trainer 9(1),* 1-19.

Silva, C. & Moran, S. (2022). *Creophonpt 1.0.* DOI: 10.5281/ZENODO.7575862. https://zenodo.org/record/7575862.

Trigo, L., Silva, C., & Moitinho, V. (2023). Learning Phonology with data in the classroom: engaging students in the creolistic research process. In *HUGOD Conference – Humanities Going Digital: Teaching, training & research experiences*, Lisbon. Nova University. Available at: https://hugod.fcsh.unl.pt.

Wichmann, S., & Holman, E. (2009). *Temporal stability of linguistic typological features.* Lincom Europa.