

INTEGRAÇÃO EM PROJETO DE INVESTIGAÇÃO  
PARTICIPATION IN RESEARCH PROJECTS  
2º SEMESTRE 2023/2024

1. Unidade curricular/ <i>Course unit</i> :	
a) Designação: INICIAÇÃO À INVESTIGAÇÃO EM LINGÜÍSTICA DE CORPUS E EXTRAÇÃO DE INFORMAÇÃO	
Name: INTRODUCTION TO RESEARCH IN CORPUS LINGUISTICS AND INFORMATION EXTRACTION	
b) Número de vagas/ <i>vacancies</i> : 2	
2. Unidade de Investigação/ <i>Research Unit</i> : Centro de Linguística da Universidade NOVA de Lisboa / Linguistics Research Center of NOVA University Lisbon	
3. Curso/ <i>Course</i> : <b>Opção livre</b> aberta aos cursos de: (indicar um dos cursos)	
Licenciatura      Mestrado      X      Doutoramento	
4. Ano do plano de estudos/ <i>Syllabus year</i> : 2023	
5. Número de créditos/ <i>Number of credits (1 ECTS 28 horas)</i> :	
Licenciatura 6 ECTS      Mestrado 10 ECTS      X      Doutoramento 10 ECTS	
6. Docente ou Investigador responsável/ <i>Teacher or principal Researcher</i> : Raquel Amaro )	
7. Período de funcionamento/ <i>Class period</i> : 14/02/2024 a 01/06/2024	
8. Objetivos da unidade curricular/ <i>Learning objectives</i> (máx. 200 palavras; expostos em termos do que se espera que o aluno adquira:	
1. Desenvolver o conhecimento de conceitos teóricos na área da Linguística de Corpus para extração de informação./Develop knowledge on theoretical concepts in the field of Corpus Linguistics for information extraction	
2. Desenvolver o conhecimento de desenho e planificação de um projeto de investigação, incluindo questões éticas relacionadas com processamento de dados./Develop knowledge on designing and planning a research project, including ethical issues related to data processing.	
3. Desenvolver o conhecimento de metodologias e ferramentas de Linguística de Corpus, orientadas para a extração de resultados específicos./Develop knowledge on Corpus Linguistics methods and tools, aiming at the extraction of specific results.	
4. Desenvolver a capacidade de comunicação de ciência, nomeadamente escrita de artigos e apresentação de trabalho./Develop science communication skills, namely paper writing and work presentation.	
9. Requisitos de frequência/ <i>Attendance requirements</i> : não se aplica/it does not apply.	
10. Conteúdo da unidade curricular/ <i>Syllabus</i> (máx. 200 palavras): O projeto EXPRESSO - Exploração preliminar de corpus para ensino inclusivo da língua de acolhimento baseado em tarefas - tem como objetivo a investigação preliminar sobre o ensino-aprendizagem do português como língua de acolhimento, com base na exploração de textos informativos e institucionais autênticos direcionados a pessoas migrantes. Esta investigação inclui a exploração de i) métodos de identificação, extração e	

descrição de vocabulário fundamental para a aprendizagem de língua de acolhimento, utilizando ferramentas e procedimentos computacionais; ii) análise semiautomática de textos e seus respetivos contextos com vista à identificação e extração de traços e critérios para monitorização de tendências/preconceitos nos textos e no uso da língua.

Como resultado geral, o projeto visa informar e sustentar o desenho da uma implementação a uma maior escala. Especificamente, são esperados como resultados um corpus para fins específicos com foco na integração e acolhimento de pessoas migrantes; um léxico de frequências para fins específicos (integração pela língua); e análise de tendências/marcadores de viés cultural e exclusão, incluindo análise preliminar dos traços (features) relevantes com base em dados quantitativos e qualitativos do corpus./ The project EXPRIMI - Preliminary corpus exploration for inclusive task-based teaching of the host language - aims to carry out preliminary research into the teaching-learning of Portuguese as a host language, based on the exploration of authentic informative and institutional texts directed to migrant people. This research includes the exploration of i) methods for identifying, extracting and describing fundamental vocabulary for learning a host language, using computational tools and procedures; ii) semi-automatic analysis of texts and their respective contexts with a view to identifying and extracting features and criteria for monitoring trends/bias in texts and language use.

As a general result, the project aims to inform and support the design of a larger-scale implementation. Specifically, the expected results are a corpus for specific purposes focusing on the integration and reception of migrants; a frequency lexicon for specific purposes (language integration); and analysis of trends/markers of cultural bias and exclusion, including preliminary analysis of relevant features based on quantitative and qualitative data of the corpus.

**11. Bibliografia recomendada/Recommended reading:** (máx. 5 títulos. Por ordem decrescente de data de edição.)

1. Amaro, R., S. Correia, M. Gonçalves, C. Barbero & M. Magalhães (2022). Inclusive Host Language Teaching: a case study of Portuguese as Foreign Language materials targeting migrants and refugees. In: Meletiadou, E. (ed.) Policies and Practices for Assessing Inclusive Teaching and Learning. United Kingdom: IGI Global, pp. 183-210. United Kingdom: IGI Global, 2022. ISBN13: 9781799885795
2. Pustejovsky, J. , Harry, B. and Zaenen. A. (2017). Designing Annotation Schemes: From Theory to Model. In: Pustejovsk, P. and, Ide, N. (eds.) Handbook of Linguistic Annotation. Springer. DOI 10.1007/978-94-024-0881-2, 21-71
3. Finlayson, M. A. and Erjavec, T. (2017). Overview of Annotation Creation: Processes and Tools. In: Pustejovsk, P. and, Ide, N. (eds.) Handbook of Linguistic Annotation. Springer. DOI 10.1007/978-94-024-0881-2, 167-191.
4. McEnery, T. & A. Hardie (2012). Corpus Linguistics: Method, theory and practice. Cambridge University Press.
5. Ben-Dov, M., Feldman, R. (2005). Text Mining and Information Extraction. In: Maimon, O., Rokach, L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. [https://doi.org/10.1007/0-387-25465-X\\_38](https://doi.org/10.1007/0-387-25465-X_38).

12. **Métodos de ensino/Teaching Methods:** Formação inicial sobre a área científica, o projeto de investigação e as tarefas a desempenhar. O/a estudante será integrado/a na equipa e o seu trabalho será acompanhado pela equipa de investigação em sessões de orientação tutorial./Initial training on the scientific field, the research project and the tasks to perform. The student will be integrated in the team and her/his work will be supervised by the research team in tutorial sessions.

13. **Métodos de avaliação/Assessment methods.** (para cada elemento de avaliação deve indicar a respetiva ponderação): Avaliação contínua, incluindo da capacidade de execução das tarefas atribuídas, da capacidade de contribuir para o trabalho em equipa e da produção de textos científicos (relatório ou artigo)./Continuous evaluation, including the ability for performing the assigned tasks, the ability to contribute for the team work and the production of scientific texts (report or paper).

14. **Língua de ensino/Teaching language:** Português ou Inglês/Portuguese or English