

Titulo: A educação e a língua nas memórias do trabalho

## Descrição do Projecto

O projecto engloba investigadores e estudantes de diferentes áreas das ciências sociais (Linguística, História Social, Ciências da Educação e Ciências da Documentação) e da Ciência de Computadores e tem como objectivo contribuir, numa perspectiva multi-disciplinar, para o estudo das condições de trabalho e formação dos trabalhadores no Porto, na 2ª metade do século XX, bem como dos processos de construção identitários, pela análise de narrativas biográficas.

O projecto tem como base e prossegue trabalho já desenvolvido pelo Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto da Universidade Popular do Porto (CDI) - <http://cdi.upp.pt>, com o qual estes investigadores têm colaborado, e que tem como objectivo a preservação e divulgação da memória e da história oral e social do Porto do século XX. No âmbito da história oral foram já realizadas 80 entrevistas, com duração superior a 260 horas, a trabalhadores de diferentes profissões e com diferentes experiências de intervenção social. A qualidade da informação recolhida impõe-se como material de estudo em diferentes áreas científicas.

Este projecto tem dois eixos de investigação transdisciplinares com vista à caracterização dos processos de inter-estruturação entre trabalho, formação e identidade pessoal, um que foca uma análise linguística e socio-linguística e outro uma análise dos processos de educação e formação, ambos interligados com a investigação no processamento computacional da língua portuguesa e no processamento de documentos estruturados.

Pretende-se contribuir para a análise diversas características linguísticas do corpus das narrativas e a sua relação com as condições sociais, de formação, etárias e dialectais dos intervenientes. Do ponto de vista linguístico, um factor relevante é o da oralidade, uma vez que o corpus é constituído por transcrições de entrevistas cujo sonoro e e imagem está acessível. A identificação de variedades sincrónicas (regionais e sociais) associadas ao corpus oral a analisar, bem como a constatação de mudanças diacrónicas no léxico e no uso de determinadas construções sintácticas, são exemplos de resultados esperados neste projecto.

Inicialmente será feita uma selecção de alguns itens lexicais e padrões de construções sintácticas a pesquisar e a identificar computacionalmente, assim como serão desenvolvidas as aplicações informáticas necessárias. Seguidamente, proceder-se-á à análise estatística e à construção de uma base de conhecimento de padrões relevantes. Finalmente será feita a avaliação dos resultados com vista a uma possível abordagem sócio-linguística, nomeadamente, marcas regionais, de poder e de género.

Do ponto de vista das ciências da educação a preocupação incidirá na compreensão do modo como nas trajetórias biográficas dos trabalhadores se articulam processos de educação e de formação que transcendem os espaços, os tempos e os actores institucionalmente definidos como educativos, processos inseparáveis da construção de uma

identidade operária e onde as dimensões da educação profissional, cívica e política se encontram articuladas.

Para a organização e categorização das narrativas serão usadas aplicações informáticas de processamento de documentos estruturados, já desenvolvidas e em desenvolvimento por membros da equipa de investigação.

## Plano de trabalho

### Objectivos/Resultados esperados

Relativamente ao conjunto da equipa

1. Contribuir para a consolidação duma equipa multidisciplinar que inclua investigadores de diferentes áreas científicas que desenvolvam trabalho junto de jovens investigadores.
2. Contribuir para a compreensão dos processos de interação entre trabalho, formação pessoal e identidade profissional.
3. Contribuir para o processamento computacional de língua portuguesa e o desenvolvimento de aplicações informáticas em software livre que permitam a manipulação de documentos estruturados que facilitem o seu acesso, classificação e pesquisa.

Relativamente ao trabalho com os estudantes na área da linguística e do processamento computacional da língua portuguesa:

1. Identificação e descrição de padrões linguísticos utilizados pelos intervenientes das narrativas.
2. Análise de variações lexicais e de mudanças no uso de certas construções sintáticas numa perspectiva sócio-linguística.
3. Desenvolvimento de aplicações informáticas de uso genérico para a análise léxica e sintática de corpora, assim como de análises estatísticas e identificação de padrões de construções sintáticas.

Relativamente ao trabalho com os estudantes na área das ciências da educação e na processamento computacional de documentos estruturados:

1. Análise de um conjunto de narrativas biográficas de trabalhadores e activistas sociais visando esclarecer a dimensão educativa associada aos espaços de sociabilidade e de luta política dos trabalhadores na segunda metade do século XX na cidade do Porto.
2. Identificar e caracterizar os processos de autoformação e de aprendizagem experiencial envolvidos no exercício profissional
3. Problematicar a dimensão da educação cívica e política envolvida nos processos de trabalho e de companheirismo no trabalho.

## Tarefas

I A realizar pela equipa de investigadores com os estudantes na área da linguística e do processamento computacional da língua portuguesa:

1.
  - 1.1. Selecção de itens lexicais e de construções sintácticas a pesquisar.
  - 1.2. Análise, integração e construção de aplicações informáticas para acesso a dicionários, etiquetação de corpora e análise morfológica.
2.
  - 2.1. Desenho e implementação duma aplicação para a diversas análises estatísticas de léxico, entre elas:
    - ocorrência de vocábulos
    - ocorrência de categorias sintácticas
  - 2.2. Avaliação dos resultados estatísticos e da sua relevância para a abordagem sócio-linguística do corpus.
3.
  - 3.1. Identificação de padrões de construções sintáticas, nomeadamente de:
    - casos de subordinação (temporal, concessiva e completiva);
    - Construções passivas.
  - 3.2. Desenvolvimento de uma aplicação para a detecção e filtragem das construções sintáticas identificadas.
  - 3.4. Análise dos resultados e avaliação das aplicações informáticas desenvolvidas
4.
  - 4.1. Identificação de outros padrões linguísticos interessantes
  - 4.2. Construção duma base de conhecimento -- tree-bank-- com os padrões encontrados e desenvolvimento de aplicações informáticas para sua manipulação.
5.
  - 5.1. Avaliação e análise socio-linguística dos resultados
  - 5.2. Avaliação das aplicações informáticas desenvolvidas e sua disponibilização em software livre.
  - 5.3. Preparação da apresentação pública do projecto

II A realizar pela equipa de investigadores com os estudantes na área das ciências da educação e na processamento computacional de documentos estruturados:

1. Organização e tratamento dos documentos que constituem as narrativas;
2. Leitura "flutuante" das narrativas biográficas recolhidas;
3. Construção de documentos estruturados associados às narrativas, utilizando aplicações informáticas desenvolvidas no âmbito do projecto.
4. Identificação de temáticas constantes das narrativas.
5. Construção de categorias emergentes das narrativas.
6. Construção de redes de relações inter-categoriais.
7. Classificação e anotação das narrativas utilizando aplicações informáticas para a classificação de documentos estruturados, em desenvolvimento no projecto.

8. Análise e interpretação teórica das categorias emergentes e pré-definidas.

9. Produção de relatório e preparação da apresentação pública do projecto.

#### Cronograma

1. mês I.1.1, I.1.2, II.1
2. mês I.2.1 I.2.2, II.2
3. mês I.2.2, II.2, II.3
4. mês I.3.1, II.4, II.5
5. mês I.3.2, II.5, II.6
6. mês I.3.4, I.4.1, II.7
7. mês I.4.1, I.4.2, II.8
8. mês I.5.1, I.5.2, II.8
9. mês I.5.3, II.8, II.9